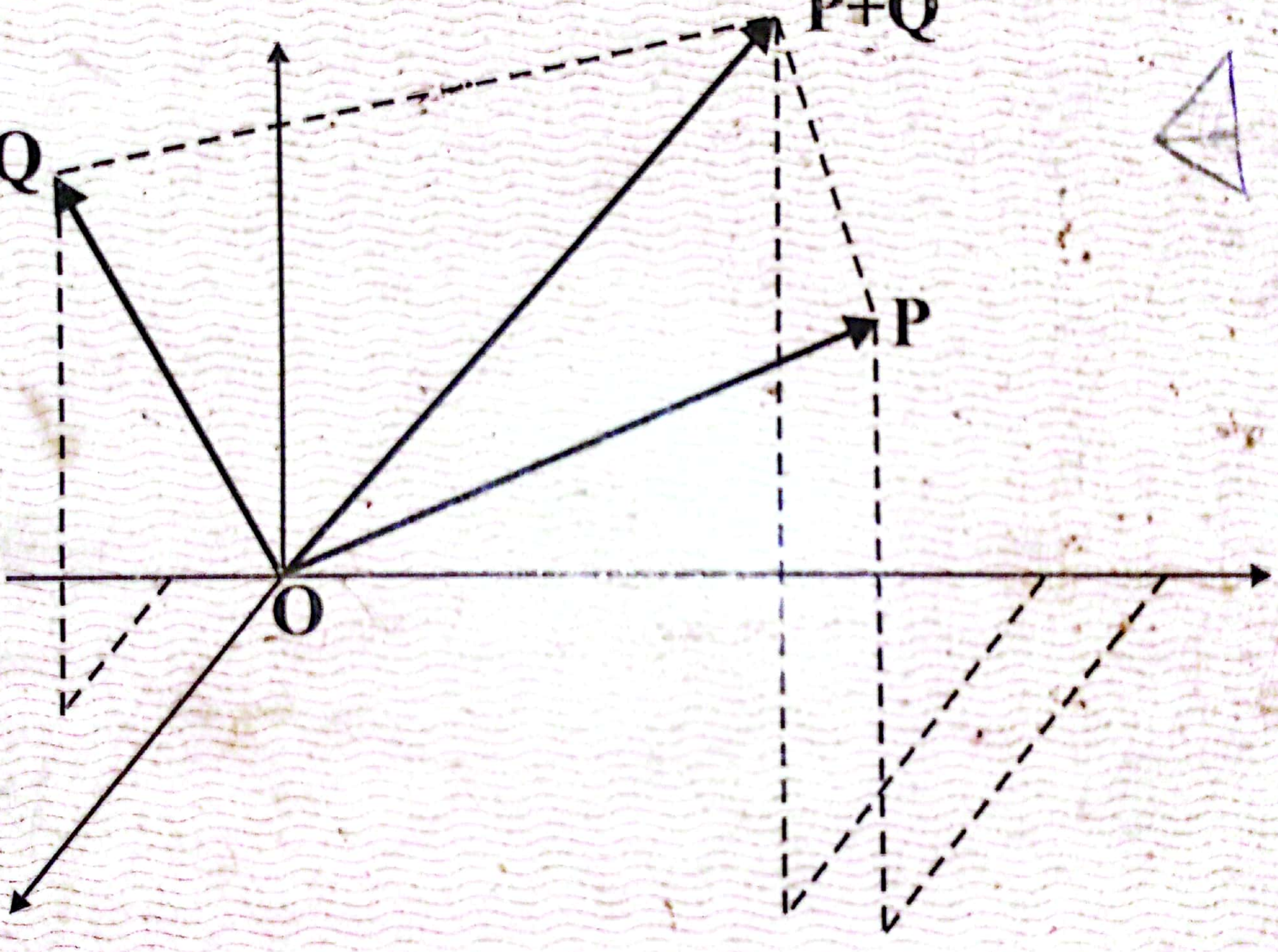
UNDERGRADUATE MATHEMATICAL TEXT

INTRODUCTION
TO
LINEAR ALGEBRA

CH



S. OMOLOYE AJALA

CONTENTS

Chapter One: Vectors 1

1.1 Addition and scalar multiplication of vectors 1

1.2.1 Dot product and norm of vectors in 4

1.3 Complex numbers........8

1.4 Exercise 10

Chapter Two: Vector spaces 12

2.1 Definition and examples........12

2.2 Subspaces........16

2.3 Basis and dimension........19

2.4 Sums and direct sums 27

2.5 Exercise 29

Chapter Three: Matrices 31

3.1 Definition and examples........31

3.1 Equality of matrices, addition and scalar multiplication........ 33

3.3 Multiplication of matrices........34

3.4 Transpose, row equivalence and elementary row operations........ 36

3.5 Applications........41

3.6 Exercises........45

Chapter Four: Linear mappings 47

4.1 Mapping........47

4.2 Linear mapping........49

4.3 Matrices and linear operators........53

4.4 Trace........64

Exercise 4.5........66

Chapter Five: Determinants 67

5.1 Definition and examples 67

5.2 Properties of determinants and determinant of matrices of large size 70

5.3 Application of determinants 73

5.4 linear equations 80

5.5 Exercise 89

Chapter Six: Eigenvalues and eigenvectors 91

6.2 Characteristics polynomial 92

6.3 Minimal polynomial 98

6.4 Exercise 103

Chapter Seven: Linear functional and dual 104

7.1 Definition and examples 104
7.2 Dual basis 106
7.3 Annihilators and transpose of a linear map 111
7.4 Exercise 113

Chapter Eight: Bilinear forms 115

8.1 Definition and examples 115
8.2 Skew-symmetric and symmetric bilinear forms 120
8.3 Quadratic forms 127
8.4 Hermitian forms 130
8.5 Exercises 132

Chapter Nine: Inner product spaces 135

9.1 Definition and examples 135
9.2 Orthogonality and orthonormal vectors 137
9.3 Linear functionals and adjoint linear operators 142
9.4 Orthogonal and unitary operators 146
9.5 Exercise 155
Bibliography 157

CHAPTER ONE

VECTORS

1.1 Addition and Scalar Multiplication of Vectors

110313091

In many scientific applications, certain quantities only have magnitudes. These quantities are referred to as scalars. For example age, height, temperature and speed are examples of scalars. There are other quantities which have both magnitudes and direction. Some examples of such quantities are velocity, acceleration and force. These are called vectors. It therefore follows that the essential distinction between a vector quantity and a scalar quantity is that a vector quantity possesses direction as well as magnitude, whereas scalar quantity possesses magnitude only.

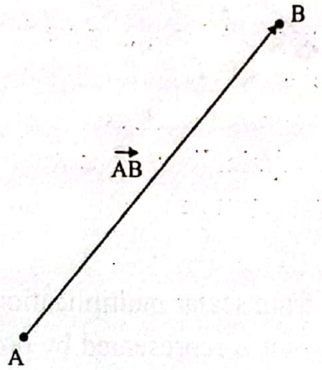


Figure 1

A vector can be denoted by letters in order of its direction, written with an arrow on top. For example, $\vec{AB}$ is a vector originating from point A and in the direction of point B.

We can also denote a vector by a single small letter with an arrow on top, $\vec{a}$, say. The magnitude of the vector represented by the distance AB is always positive and can be written as $|\vec{AB}|$ or $|\vec{a}|$. It is usually called the modulus of the vector.

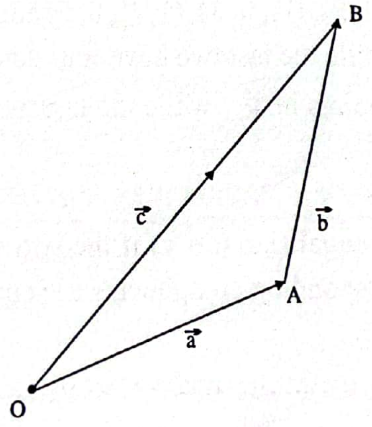


Figure 2

If a particle is at a point O and it is displaced to point A and another displacement which takes it to point B. The effect of these two displacements is identical to the displacement from O to B. If we denote $\vec{OA}$, $\vec{AB}$ and $\vec{OB}$ by $\vec{a}$, $\vec{b}$ and $\vec{c}$, see Figure 2, then we say that $\bar{c}$ is the vector sum of $\bar{a}$ and $\bar{b}$ and we write $\bar{a}+\bar{b}=\bar{c}$. If we apply a minus sign to a vector, it means that the direction of the vector is reversed hence we interprete $\bar{a}-\bar{b}=\bar{a}+(-\bar{b})$.

If we multiply a vector by a scalar, the modulus of the vector is altered. If k is a scalar and $\bar{a}$ is a vector then $k\bar{a}$ as the same direction as $\bar{a}$ but its modulus is $|k\bar{a}|=k|\bar{a}|=k|\bar{OA}|=|O'A|$, see Figure 4.

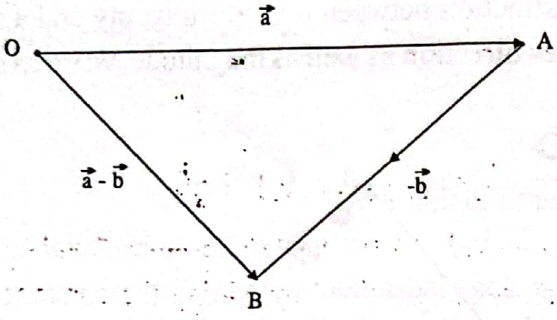


Figure 3

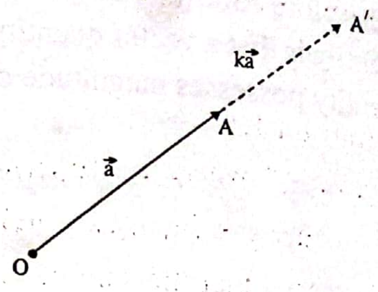


Figure 4

This procedure is different from scalar multiplication of two vectors.
Recall that in the plane, a point is represented by ordered pair of real numbers. Thus an ordered pair (a, b) of real numbers is a vector. We generalize this notion.

Definition 1.1.1

The set of all n-tuples of real numbers, denoted by R^n is called an n-space, an n-tuple $v=(v_1, v_2, \ldots, v_n)$ is called a point or vector in R^n. The real numbers v_i are called the components (or coordinates) of the vector v.

Example 1.1.1

Consider the vectors: (3, 1, 2), (-1, 0, 3), (1, -1, 0, 5) and (0, -3, 4, 0). The first two vectors have three components while the last two have four components. It therefore follows that the first two vectors are points in R^3 while the last two are in R^4.

Definition 1.1.2

Two vectors u and v are equal (i.e. u = v) if the two vectors have the same number of components and the corresponding components are equal.

If $u, v \varepsilon R^n$ where $u=u_1, u_2, \ldots, u_n$ and $v=v_1, v_2, \ldots, v_n$ then

$$u=v$$

iff $u_i=v_i$ for $1 \leq i \leq n$. Note that the vectors (3,8,-1) is not equal to (8,3,-1).

Example 1.1.2.

If u = (8, 1, 6) and v = (a + b, c - 1, a - b) are two vectors in R^3 find a, b and c if u=v.

Solution

Since u = v then

$$(8, 1, 6) = (a + b, c - 1, a - b)$$

and so,

$$a + b = 8 \quad (1)$$
$$c - 1 = 1 \quad (2)$$
$$a - b = 6 \quad (3)$$

From equation (2), we have

$$c = 2$$

and from equations 1 and 3 we have

$$a = 7 \text{ and } b = 1.$$

Definition 1.1.3.

If $u, v \varepsilon R^n$ where $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$ then

$$u + v = (u_1, u_2, \ldots, u_n) + (v, v_2, \ldots, v_n) = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n).$$

This means that the sum of the two vectors are obtained by adding the corresponding components. Notice that u + v is also a vector in R^n.

If R is a real number and u is a vector in R^n then the product of k and u denoted by ku is the vector obtained by multiplying each component of u by k i.e.

$$ku = k(u_1, \ldots, u_n) = (ku_1, ku_2, \ldots, ku_n).$$

Notice that ku is also a vector in R^n. We also have

$$-u = -1(u) \text{ and } u - v = u + (-v).$$

Example 1.1.3.

If u = (-2, 5, 7) and v = (3, 11, 4) then

$$u + v = (-2 + 3, 5 + 11, 7 + 4) = (1, 16, 11)$$
$$7u = (7 \times (-2), 7 \times 5, 7 \times 7) = (-14, 35, 49)$$
$$3u - 4v = (-6 - 12, 15 - 44, 21 - 16) = (-18, -29, 5)$$

Notice that the vector (0, 0, ..., 0) εR^n where every component is 0 is called the zero vector. The zero vector is usually denoted by 0 and it similar to the scalar 0 because

$$u + 0 = (u_1 + 0, u_2 + 0, \ldots, u_n + 0) = (u_1, u_2, \ldots, u_n) = u.$$

If u, v, w are vectors in R^n and a, b are scalars in R, the following are properties of the vectors in R^n

(i) (u + v) + w = u + (v + w)
(ii) u + 0 = u
(iii) u + (-u) = 0
(iv) u + v = v + u
(v) a(u + v) = au + av
(vi) (a + b)u = au + bu
(vii) (ab)u = a(bu)
(viii) 1u = u.

These properties of R^n make R^n to be a vector space over R. The idea of a vector space will be presented in the next chapter and the above properties of R^n will be proved in the next chapter.

1.2.1 Dot Product and Norm of Vectors in R^n

In R^n we are usually interested in angle between two given vectors. We need to know if two vectors are orthogonal or they are parallel. The length of a given vector is also of interest in scientific and engineering investigations.

Definitions 1.2.1.

Let $u, v \varepsilon R^n$ where $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$ the dot product (or inner product, or scalar product) denoted by u.v is the scalar obtained by multiplying corresponding components and adding the results. That is,

$$u.v = u_1 \; v_1 + u_2 \; v_2 + \ldots + u_n \; v_n$$

If the dot product of u and v is zero then we say that the two vectors are orthogonal or are perpendicular.

Example 1.2.1

Let $u, v, w \varepsilon R^5$ where $u = (5, -8, 5, 7, -2)$, $v=(1, -2, 3, -4, 4)$ and $w = (3, 1, 7, -3, 1)$, find $u \bullet v$ and $u \bullet w$.

Solution

$$u.v = (5, -8, 5, 7, -2).(1, -2, 3, -4, 4)$$
$$= 5 \times 1 + (-8)(-2) + 5 \times 3 + 7 \times (-4) + (-2) \times 4 \quad 5 + 16 + 15 - 28 - 8 = 0$$
$$u.w = (5, -8, 5, 7, -2).(3,1,7,-3,1)$$
$$= 5 \times 3 + (-8) \times 1 + 5 \times 7 + 7 \times (-3) + (-2) \times 1 \quad 15 - 8 + 35 - 21 - 2 = 19$$

This shows that vectors u and v are orthogonal vectors.

Example 1.2.2

If $u, v, w \varepsilon R^n$ and $c \varepsilon R$ then show the following

(i) $(u + v) \bullet w = u \bullet w + v \bullet w$

(ii) $(cu) \bullet v = c(u \bullet v)$

(iii) $u \bullet v = v \bullet u$

(iv) $u \bullet u = 0$ and $u \bullet u = 0$ if and only if $u = 0$

Solution

Since $u, v, w \varepsilon R^n$ then the vectors must have n-components hence let $u = (u_1, u_2, \ldots, u_n)$, $v = (v_1, v_2, \ldots, v_n)$ and $w = (w_1, w_2, \ldots, w_n)$.

(i) $(u + v).w = \{(u_1, u_2, \ldots, u_n) + (v_1, v_2, \ldots, v_n)\} \bullet (w_1, w_2, \ldots, w_n)$

$$= (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n) \bullet (w_1, w_2, \ldots, w_n)$$
$$= (u_1 + v_1) \; w_1 + (u_2 + v_2) \; w_2 + \ldots + (u_n + v_n) \; w_n$$
$$= u_1 \; w_1 + v_1 \; w_1 + u_2 \; w_2 + u_2 \; w_2 + \ldots + u_n \; w_n + v_n \; w_n$$
$$= u_1 \; w_1 + u_2 \; w_2 + \ldots + u_n \; w_n + v_1 \; w_1 + v_2 \; w_2 + \ldots + v_n \; w_n$$
$$= u.w + v.w.$$

(ii) $cu = c(u_1, u_2 \ldots, u_n) = (cu_1, cu_2, \ldots, cu_n)$

$$\therefore (cu).v = (cu_1, cu_2, \ldots, cu_n).(v_1, v_2, \ldots, v_n) = cu_1 v_1 + cu_2 v_2 + \ldots + cu_n v_n$$

$$= c(u_1 v_1 + u_2 v_2 + \ldots + u_n v_n) = c(u.v)$$

(iii) $u.v = (u_1, u_2, \ldots, u_n)(v_1, v_2, \ldots, v_n) = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$

Since u_i and v_i are elements in R then

$$u_i v_i = v_i u_i$$

hence the above equals

$$v_1 u_1 + v_2 u_2 + \ldots + v_n u_n = v.u.$$

(iv) $u.u = (u_1, u_2, \ldots, u_n).(u_1, u_2, \ldots, u_n)$

$$= u_1 u_1 + u_2 u_2 + \ldots + u_n u_n = u_1^2 + u_2^2 + \ldots + u_n^2$$

since each u_i^2 is non-negative for all $i = 1, 2, \ldots, n$ and sum of non-negative real numbers is non-negative then

$$u.u = u_1^2 + u_2^2 + \ldots + u_n^2$$

is non-negative.

$$\therefore u.u \geq 0.$$

$u.u = 0$ if and only if $u_i^2 = 0$, i.e. if and only if $u_i = 0$ for each $i = 1, 2, \ldots n$ i.e. if and only if $u = 0$.

Definition 1.2.2

Let u and v be vectors in R^n, the distance between u and v denoted by $|u - v|$ is defined by

$$|u - v| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \ldots + (u_n - v_n)^2}$$

Recall that in example 1.2.2, $u.u \geq 0$, in fact

$$u.u = u_1^2 + u_2^2 + \ldots + u_n^2 \geq 0$$

therefore

$$\sqrt{u.u} = \sqrt{u_1^2 + u_2^2 + \ldots + u_n^2}.$$

We denote $\sqrt{u.u}$ by $||u||$ and call it the norm or the length of u. Note that the square root exist because $u.u \geq 0$.

Example 1.2.3

Let $u = (2, 3, -1, 0)$ and $v = (4, 6, 5, 1)$, then

$$||u - v|| = \sqrt{(2-4)^2 + (3-6)^2 + (-1-5)^2 + (0-1)^2}$$

$$= \sqrt{4+9+36+1} = \sqrt{50} = 5\sqrt{2}$$

$$||u|| = \sqrt{2^2 + 3^2 + (-1)^2 + 0^2} = \sqrt{4+9+1+0} = \sqrt{14}$$

$$||v|| = \sqrt{4^2 + 6^2 + 5^2 + 1^2} = \sqrt{17+36+25+1} = \sqrt{78}$$

Definition 1.2.3

A vector $u \varepsilon R^n$ is called a unit vector if its norm is 1 i.e. $||u|| = 1$.

Notice that any non-zero vector $w \varepsilon R^n$ can be reduced to a unit vector $e = \dfrac{w}{\|w\|}$.

For $||e|| = \left|\dfrac{w}{\|w\|}\right| = \dfrac{\|w\|}{\|w\|} = 1.$

Theorem 1.2.1

Let $u, v \varepsilon R^n$ then $|u.v| \leq ||u||\ ||v||$.

Proof:

If $u = 0$ and $v \neq 0$ then $|u.v.| = 0$ and $||u||\ ||v|| = 0$ and so the result follows.;
Similarly, if $v = 0$ and $u \neq 0$ the result reduces to $0 \leq 0$ and so the result is true. We only need to consider the case in which $u \neq 0$ and $v \neq 0$, then

$$||u|| \neq 0 \text{ and } ||v|| \neq 0.$$

If $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$

$$|u.v| = |\ u_1\ v_1 + u_2\ v_2 + \ldots + u_n\ v_n\ |$$
$$\leq |\ u_1\ v_1| + |\ u_2\ v_2\ | + \ldots + |u_n\ v_2| = \Sigma\ |u_i\ v_i\ |$$

Notice that if a, b are real numbers, then $(a-b)^2 \geq 0$ thus

$$0 \leq (a-b)^2 = a^2 - 2ab + b^2$$
$$\therefore 0 \leq a^2 - 2ab + b^2$$

implies

$$2ab \leq a^2 + b^2.$$

Let $a = \dfrac{|u_i|}{\|u\|}$ and $b = \dfrac{|v_i|}{\|v\|}$, then on substituting for a and b in the above we have

$$2\frac{|u_i|}{\|u\|}\frac{|v_i|}{\|v\|} \leq \frac{|u_i|^2}{\|u\|^2} + \frac{|v_i|^2}{\|v\|^2}$$

we take the summation to obtain

$$2\frac{\Sigma|u_i|v_i|}{\|u\|\|v\|} \leq \frac{\Sigma\ |u_i|^2}{\|u\|^2} + \frac{\Sigma\ |v_i|^2}{\|v\|^2}.$$

Recall that

$$\|u\|^2 = \Sigma u_i^2 = \Sigma|u_i|^2$$

and

$$\|v\|^2 = \Sigma v_i^2 = \Sigma|v_i|^2,$$

when these values are substituted for $\Sigma|u_i|^2$ and $\Sigma|v_i|^2$ we find that

$$2\frac{\Sigma|u_i\ v_i|}{\|u\|\ \|v\|} \leq \frac{\Sigma|u_i|^2}{\|u\|^2} + \frac{\Sigma|v_i|^2}{\|u\|^2} = \frac{\|u\|^2}{\|u\|^2} + \frac{\|v\|^2}{\|u\|^2} = 1 + 1 = 2$$

and on dividing both sides by 2 we have

$$\frac{\Sigma|u_i\ v_i|}{\|u\|\ \|v\|} \leq 1$$

On multiplying both sides by $||u||\ ||v||$, we have

$$\Sigma\ |u_i\ v_i| \leq \|u\|\ \|v\|.$$

Thus we have

$$|u.v| \leq \Sigma|u_i\ v_i| \leq \|u\|\ \|v\|.$$

This theorem is sometimes called Cauchy-Schwarz inequality.

Theorem 1.2.2

If u, v ε R^n then $||u + v|| \leq ||u|| + ||v||$.

Proof:

If $||u + v || = 0$ then the inequality holds. We therefore consider the case where $||u + v || = 0$.

Let $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$. By the property of absolute value on set of real numbers and since u_i and v_i are real numbers then

$$|u_i + v_i| \leq |u_i| + |v_i|$$

$$\|u + v\|^2 = \Sigma(u_i + v_i)^2 = \Sigma|u_i + v_i|^2 = \Sigma|u_i + v_i||u_i + v_i|$$

$$\leq \Sigma|u_i + v_i|(\ |u_i| + |v_i|\ ) = \Sigma|u_i + v_i||u_i| + \Sigma|u_i + v_i|\,|v_i|$$

By the preceding Cauchy-Schwarz inequality we have

$$\Sigma|u_i + v_i\|u_i| \leq \|u + v\|\|u\|$$

and

$$\Sigma|u_i + v_i\|v_i| \leq \|u + v\|\|v\|.$$

Thus

$$\|u + v\|^2 \leq \|u + v\|\|u\| + \|u + v\|\|v\|$$

On dividing both sides of the inequality by $||u + v|| = 0$, we have

$$||u + v|| \leq ||u|| + ||v||.$$

inequality is sometimes called Minkowski inequality.

Definition 1.2.4

Let u, v ε R^n and if θ is the angle between u and v where $u = 0$ and $v = 0$ then

$$\cos\theta = \frac{u.v}{\|u\|\|v\|}.$$

Notice that if u.v. = 0 then $\cos\theta = 0$. This implies $\theta = 90°$. As we have discussed earlier, it follows that two non-zero vectors u, v ε R^n are perpendicular (or orthogonal) if and only if their scalar product u.v is zero.

Example 1.2.4

If u, v ε R^4 such that u = (1, 4, -2, 3) and v = (-3, 1, 6, 7). Find the cosine of the angle between u and v.

Solution

$$u.v = (1, 4, -2, 3).(-3, 1, 6, 7) = -3 + 4 - 12 + 21 = 10$$

$$||u|| = \sqrt{1^2 + 4^2 + (-2)^2 + 3^2} = \sqrt{1 + 16 + 4 + 9} = \sqrt{30}$$

$$||v|| = \sqrt{(-3)^2 + 1^2 + 6^2 + 7^2} = \sqrt{9 + 1 + 36 + 49} = \sqrt{95}$$

If θ is the angle between u and v,

$$\cos\theta = \frac{u.v}{\|u\|\|v\|} = \frac{10}{\sqrt{30}\sqrt{95}} = \frac{10}{\sqrt{2850}}$$

1.3 Complex Numbers

Definition 1.3.1

A complex number is a number of the form a + ib where a and b are real numbers and $i^2 = -1$. Thus a complex number a + ib consists of the sum of a real part a and the imaginary part ib.

The set of complex numbers is denoted by C. If a + ib and c + id are complex numbers then

$$(a + ib) + (c + id) = a + c + i(b + d)$$

and

$$a + ib) - (c + id) = a - c + i(b - d).$$

$$(a + ib)(c + id) = ac - bd + i(ad + bc).$$

Conjugate of a complex number z = a + ib is denoted by

$$\bar{z} = a - ib.$$

Note that the sum and product of a complex number and its conjugate are real.

$$z + \bar{z} = (a + ib) + (a - ib) = 2a$$

and

$$z\bar{z} = (a + ib)(a - ib) = a^2 - iab + iab - i^2 b^2 = a^2 + b^2.$$

If z == 0 then the inverse of z denoted by z^{-1}

$$z^{-1} = \frac{1}{z} = \frac{1}{a + ib} = \frac{a - ib}{(a + ib)(a - ib)} = \frac{a - ib}{a^2 + b^2} = \frac{a}{a^2 + b^2} - i\frac{b}{a^2 + b^2}$$

hence z^{-1} is also a complex number.

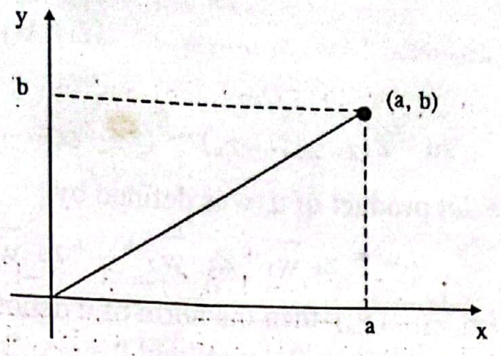


Figure 5

Recall that any real number can be represented by a point on a real line, in similar manner, a complex number can be represented by a point on the plane. For example a point (a, b) ε R^n represents the complex number z = a + ib.

The x-axis represent the real parts of the complex numbers while the y-axis represent the imaginary parts of the complex numbers.

The absolute value $|z| = |a + ib| = \sqrt{a^2 + b^2}$ is the distance from z to the origin. We can associate z ε C as a vector in R^2 and |z| as the norm of a vector (a, b) in R^2.

If z = a + ib then $\bar{z}$ = a - ib then

$$|z| = \sqrt{a^2 + b^2}$$

and

$$z\bar{z} = (a + ib)(a - ib) = a^2 + b^2.$$

hence

$$|z| = \sqrt{a^2 + b^2} = (z\bar{z})^{\frac{1}{2}}.$$

Example 1.3.1

Let z = 5 + i6 and w = 12 + i9, then

$$|z| = \sqrt{5^2 + 6^2} = \sqrt{25 + 36} = \sqrt{61}$$

$$|w| = \sqrt{12^2 + 9^2} = \sqrt{144 + 81} = \sqrt{225} = 15$$

$$w^{-1} = \frac{1}{12 + i9} = \frac{12 - i9}{(12 + i9)(12 - i9)} = \frac{12}{225} - \frac{i9}{225} = \frac{4}{75} - \frac{i3}{75}.$$

Definition 1.3.2

A complex n-space is the set of all n-tuples of complex numbers and it is denoted by C^n. Just as in R^n, the elements in C^n are called points or vectors and the elements in C are the scalars.

If u, w ε C^n then

$$u = (z_1, z_2, z_3, \ldots, z_n)$$

and

$$w = (w_1, w_2, \ldots, w_n)$$

where $z_i, w_i \,\varepsilon\, C$.

$$u + w = (z_1, z_2, \ldots, z_n) + (w_1, w_2, \ldots, w_n) = (z_1 + w_1, z_2 + w_2, \ldots, z_n + w_n)$$

If $z \varepsilon C$ then

$$zu = z(z_1, z_2, \ldots, z_n) = (zz_1, zz_2, \ldots, zz_n)$$

If u, w ε C^n, then the dot product of u, w is defined by

$$u.w = z_1 \overline{w_1} + z_2 \overline{w_2} + \ldots + z_n \overline{w_n} \cdot$$

If u ε C^n where $u = (z_1, z_2, \ldots, z_n)$ then the norm of u denoted by

$$|u| = \sqrt{u.u} = (z_1 \cdot \overline{z_1} + z_2\ z_2 + \ldots + z_n\ z_n)^{\frac{1}{2}}$$

since $z_i\ \overline{z_i} = |z_i|^2$ then

$$||u|| = \sqrt{u.u} = \sqrt{|z_1|^2 + |z_2|^2 + \ldots + |z_n|^2}$$

since $|z_i|^2 \geq 0$ for all $i = 1, \ldots, n$ then

$$||u|| \geq 0$$

and $||u|| = 0$ only when $u = 0$.

Example 1.3.2

Let u, v ε C^4 be such that $u = (2 - i, 2 + 3i, 4 + i, 2i)$ and $v = (5, 4 + 6i, 2 + i, 3 - i)$, then

$$\begin{aligned} u.v &= (2 - i)\overline{5} + (2 + 3i)\overline{(4 + 6i)} + (4 + i)\overline{(2 + i)} + 2i\overline{(3 - i)} \\ &= (2 - i)5 + (2 + 3i)(4 - 6i) + (4 - i)(2 - i) + 21(3 + i) \\ &= w - 5i + 8 - 12i + 18 + 8 - 4i + 2i + 1 + 6i - 2 \\ &= w + 8 + 18 + 8 + 1 - 2 + (-5i - 12i + 12i - 4i + 2i + 6i) = 43 - i \end{aligned}$$

$$\begin{aligned} u.u &= (2 - i)\overline{(2 - i)} + (2 + 3i)\overline{(2 + 3i)} + (4 + i)\overline{(4 + i)} + 2i\ \overline{(2i)} \\ &= (2 - i)(2 + i) + (2 + 3i)(2 - 3i) + (4 + i)(4 - i) + 2i(-2i) \\ &= 4 + 1 + 4 + 9 + 16 + 1 + 4 = 39 \end{aligned}$$

$$\therefore ||u|| = \sqrt{u.u} = \sqrt{39}.$$

1.4 Exercise

1. If u = (3, -1, 2), v = (-3, 2, 1) and w = (5, 3, 4) are vectors in R^3 compute
   (i) u + v - w (ii) 4v + 5w (iii) 3u + 4u - 3 (iv) 5u + 7w (v) v - u.
2. If u = (-1, -1, 1, 3) v = (0, 2, -3, 1), find
   (i) |u -v|
   (ii) ||u|| and ||v||
   (iii) u.v.
3. Determine the scalar a such that the following pair of vector u and v are orthogonal
   (i) u = (3, 6, a) and v = (6, -4, -3)
   (ii) u = (5, 3, a, 2) and v = 1, -3, 2, a + 10).
4. Find x, y, z if (10, 1, 4) = x(2, 3, 5) + y(1, 2, 4) + z(-2, 2, 3)
5. If u = (6, 2, -1, -3) v = (0, -2, 1, 7, 3), find the distance between u and v.
6. Find the value of a if the distance between u and v is 6, where u = (2, -4, 1, a) and v = (3, -1, a + 10, -3).

7. Simplify

$(3 + 5i)(8 - 6i)$ (ii) $(7-4i)^2$

$\frac{2-9i}{3i}$ (iv) $\left(\frac{1}{4-3i}\right)^2$.

8. For z, w ε C show that

$\overline{z+w} = \bar{z}+\bar{w}$ (ii) $\overline{zw} = \bar{z}\,\bar{w}$ (iii) $\bar{z} = z$.

9. Show that for any z, w ε C, zw = 0 implies z = 0 or w = 0.

10. Show that for any complex number z

(i) $|z| = |\bar{z}|$

(ii) the real part of z is $\frac{1}{2}(z+\bar{z})$

(iii) the imaginary part of z is $\frac{z-\bar{z}}{2i}$.

11. Let u, v ε C^3 where u = (3 - 8i, 4 + i, 7 + 5i) and v = (2i, -1 + i, 3 - 2i), compute

u - v (ii) v - u

u.v, v.u (iv) $||u||$ and $||v||$

(v) $||u - v||$.

12. For any vectors u, v, w ε C^n show that

u.(v + w) = u.v + u.w (ii) (v + w).u = v.u + w.u.

13. If u, v, ε C^n show that $||u + v|| \leq ||u|| + ||v||$.

14. Let u, v ε R^3, where u = (2, a, 1) and v = (4, -3, -2), determine the value of a such that u and v are perpendicular.

15. Compute the cosine of the angle the vector u = (3, -6, 2) makes with the x-axis.

M0313091

CHAPTER TWO

abelian group

VECTOR SPACES

CAIICD - field

[illegible] (+, x) conditions

2.1 Definition and Examples

Let K be a non-empty set in which two operations, addition (+) and multiplication (×) are defined. K will be called a field if the following conditions are satisfied.

(i) If a, b ε K then

$$a + b \in K \text{ and } ab \in K.$$ — closure C

(ii) For all a, b, c ε K,

$$(a + b) + c = a + (b + c) \text{ and } (ab)c = a(bc).$$ — associative A

(iii) There exists 0 ε K and 1 ε K such that for all a ε K,

$$a + 0 = 0 + a = a$$

and

$$a = a.1 = 1.a = a$$ — identity I

(iv) If a ε K then there exists -a ε K such that

$$a + (-a) = (-a) + a = 0$$ — inverse I

and if a ≠ 0 there exists a^{-1} ε K such that

$$a.a^{-1} = a^{-1}a = 1.$$

(v) For all a, b ε K,

$$a + b = b + a \text{ and } ab = ba.$$ — commutative C

(vi) For all a, b, c ε K,

$$a.(b + c) = ab + ac.$$ — distributive D

The above imply that a field (K, +, .) is a set on which two binary operations, addition (+) and multiplication (×) are defined which satisfy the following:

(i) (K, +) form an abelian group.

(ii) $(K^*,.)$ form an abelian group where K^* denotes non-zero elements of K.

(iii) For all a, b, $^c\varepsilon$ K, $a.(b+c) = ab + ac$

Example 2.1.1

The sets Q of rational numbers, R of real numbers and ℂ of complex numbers are all fields. However, the set Z of integers is not a field because $(Z^*,.)$ is not a group for if x ε Z then $x^{-1} \notin Z$.

For instance 2 ε Z but $2^{-1} = \frac{1}{2} \notin Z$ and so 2 does not have a multiplicative inverse in Z.

Definition 2.1.1

Let K be a field and V a non-empty set in which addition and multiplication by elements of K are defined, such that to any u, v ε V, the sum u + v ε V and for any u ε V and c ε K the product cu ε V. We say that V is a vector space over the field K if the following axioms are satisfied.

(i) For $u, v, w \varepsilon V$,

$$(u+v)+w=u+(v+w)$$ A

(ii) There exists an element denoted by $0 \varepsilon V$ such that for every $u \varepsilon V$,

$$u+0=0+u=u.$$ I

(iii) For any $u \varepsilon V$ there exists $-u \varepsilon V$ such that

$$u+(-u)=0.$$ I

(iv) For all $u, v \varepsilon V$,

$$u+v=v+u.$$ C

(v) If $c \varepsilon K$ and $u, v \varepsilon V$ then

$$c(u+v)=cu+cv.$$ D (+

(vi) If $a, b \varepsilon K$ and $u \varepsilon V$ then

$$(a+b)u=au+bu$$ D (+x)

(vii) If $a, b \varepsilon K$ and $u \varepsilon V$ then

$$(ab)u=a(bu)$$ A

(viii) Let $1 \varepsilon K$ be the unit in K then for any $u \varepsilon V$,

$$1u=u.$$

Notice that the first four of the above axioms show that V with the additive operation must be an abelian group.

The last four of the axioms show the multiplicative action of the field K on set V. If V is a vector space over the field K, the elements in V are called vectors while the elements of K are called scalars. We shall consider some standard notations which will be used throughout this course. Recall that in Example 2.1.1 we discussed the fact that Q, R and C are all fields. We shall be dealing with the Cartesian products of these fields.

$$Q^n = Q \times Q \times Q \times \ldots \times Q \text{ (n times)} = \{a_1, a_2, \ldots a_n \mid a_i \varepsilon Q\}$$

$$R^n = R \times R \times R \times \ldots \times R \text{ (n times)} = \{b_1, b_2, \ldots, b_n \mid b_i \varepsilon R\}$$

$$\_^n = \_ \times \_ \times \ldots \times \_ \text{ (n times)} = \{(z_1, z_2, \ldots z_n) \mid z_i \varepsilon \_\}$$

Example 2.1.2

Let K be any field, consider $K^n = \{a_1, a_2, \ldots, a_n) \mid a_i \varepsilon K\}$, we define addition and scalar multiplication thus

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$

and

$$c(a_1, a_2, \ldots, a_n) = (ca_1, ca_2, ca_3, \ldots, ca_n)$$

where $a_i, b_i \varepsilon K$. With the two operations, K^n is a vector space over K.

Solution

Let $a=(a_1, a_2, \ldots, a_n) \varepsilon K^n$ and $b=(b_1, b_2, \ldots b_n) \varepsilon K^n$ then

$$a+b=(a_1, a_2, \ldots, a_n)+(b_1, b_2, \ldots, b_n)=(a_1+b_1, a_2+b_2, \ldots, a_n+b_n)$$

since each $a_i+b_1 \varepsilon K$ then

$$a+b=(a_1+b_1, \ldots, a_n+b_n) \varepsilon K^n.$$

If $a = (a_1, a_2, \dots, a_n) \in K^n$ and $c \in K$ then

$$ca = (ca_1, ca_2, \dots, ca_n) \in K^n$$

since each $ca_i \in K$.

The first two parts of the axiom of a vector space are satisfied. If a, b, c ε K^n where $a = (a_1 a_2, a_3, \dots a_n)$, $b = (b_1, b_2, \dots, b_n)$, $c = (c_1, c_2, \dots, c_n)$ then

$$\begin{aligned}(a+b)+c &= ((a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n)) + (c_1, c_2, \dots, c_n)\\ &= (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n) + (c_1, c_2, \dots, c_n)\\ &= (a_1 + b_1 + c_1, a_2 + b_2 + c_2, \dots, a_n + b_n + c_n)\end{aligned}$$

Similarly

$$\begin{aligned}a+(b+c) &= (a_1, a_2, \dots, a_n) + ((b_1, b_2, \dots, b_n) + (c_1, c_2, \dots, c_n))\\ &= (a_1, a_2, \dots, a_n) + ((b_1 + c_1, b_2 + c_2, \dots b_n + c_n))\\ &= (a_1 + b_1 + c_1, a_2 + b_2 + c_2, \dots, a_n + b_n + c_n)\end{aligned}$$

This shows that

$$(a + b) + c = a + (b + c),$$

then definition 2.1.1 (i) is satisfied. If 0 ε K is the additive identity in the field K then consider (0, 0, ... 0) ε K^n. If a ε K^n where $a = (a_1, a_2, \dots a_n)$ then

$$\begin{aligned}(a + 0) &= (a_1, a_2, \dots a_n) + (0, 0, \dots 0)\\ &= (a_1 + 0, a_2 + 0, \dots, a_n + 0) = (a_1, a_2, \dots, a_n) = a\end{aligned}$$

hence

$$(0, 0, \dots, 0) \in K^n,$$

which is n-tuples of zeros, is the zero vector in K^n hence axiom (ii) is satisfied.

If $a = = (a_1, a_2, \dots, a_n) \in K^n$ then

$$-a = (-a_1, -a_2, \dots -a_n) \in K^n$$

and

$$\begin{aligned}a + (-a) &= (a_1, a_2, \dots, a_n) + (-a_1, -a_2, \dots, -a_n)\\ &= (a_1 - a_1, a_2 - a_2, \dots, a_n - a_n) = (0, 0, \dots, 0)\end{aligned}$$

hence (-a) is the additive inverse of a hence (iii) is satisfied.

If a, b ε K^n then

$$a + b = (a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n)$$

Since each a_i and b_i are elements of the field K and K is a field then it is commutative with respect to addition hence

$$a_i + b_i = b_i + a_i.$$

We then have

$$(a_1 + b_1, a_2 + b_2, \dots, a_n + b_n) = (b_1 + a_1, b_2 + a_2, \dots, b_n + a_n) = b + a.$$

This gives us

$$a + b = b + a$$

for all a, b ε K^n, hence (iv) is satisfied. If c ε K and a, b ε K^n then

$$c(a+b) = c((a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n)) = c(a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$
$$= (c(a_1 + b_1), c(a_2 + b_2), \ldots, c(a_n + b_n)) = (ca_1 + cb_1, ca_2 + cb_2, \ldots, ca_n + cb_n)$$

and

$$ca + cb = c(a_1, a_2, \ldots, a_n) + c(b_1, b_2, \ldots, b_n) = (ca_1, ca_2, \ldots, ca_n) + (cb_1, cb_2, \ldots, cb_n)$$
$$= (ca_1 + cb_{1_2}, ca_2 + cb_2, \ldots, ca_n + cb_n)$$

From the above result, we see that

$$c(a+b) = ca + cb$$

Hence axiom (v) is satisfied. If c, d ε K and a ε K^n then

$$(c+d)a = (c+d)(a_1, a_2, \ldots, a_n) = ((c+d)a_1, (c+d)a_2, \ldots, (c+d)a_n)$$

but since K is a field, distribution laws are satisfied hence

$$(c+d)a_i = ca_i + da_i$$

$$(c+d)a = ((c+d)a_1, (c+d)a_2, \ldots, (c+d)a_n) = (ca_1 + da_1, ca_2 + da_2, \ldots, ca_n + da_n)$$
$$= (ca_1, ca_2, \ldots, ca_n) + (da_1, da_2, \ldots, da_n) = ca + da.$$

Hence (vi) is satisfied. Also if c, d ε K and a ε K^n then

$$(cd)a = cd(a_1, a_2, \ldots, a_n) = (cda_1, cda_1, \ldots, cda_2, \ldots, cda_n)$$

Since K is a field and c, d, a_i ε K then

$$(cd)a_i = c(da_i)$$

hence

$$(cd)a = ((cd)a_1, (cd)a_2, \ldots, (cd)a_n) = c(da_2), \ldots c(da_n) = c(da_1, da_2, \ldots, da_n) = c(da).$$

We therefore have $(cd)a = c(da)$ where c, d ε K and a ε K^n hence (ii) is satisfied.

Let 1 ε K be the multiplicative identity and a ε K^n then

$$1a = 1(a_1, a_2, \ldots, a_n) = (1a_1, 1a_2, \ldots, 1a_n) = (a_1, a_2, \ldots, a_n) = a$$

hence,

$$1a = a,$$

and so axiom (vii) is satisfied. These axioms show that K^n is a vector space over K. Notice that Q^n, R^n and _^{n} are special cases of this example.

Example 2.1.3

Let K be a field and consider the set K{x} of all polynomials in x with coefficients in K. If f(x) ε K{x} then $f(x) = a_0 + a_1 x + \ldots, + a_n x^n$, where a_i ε K.

We define addition and scalar multiplication as follows: let f(x), g(x) ε F{x} then

$$f(x) = a_0 + a\ x + a_2 x^2 + \ldots + a_m x^m$$

and

$$g(x) = b_0 + b_1 x + \ldots + b_n x^n \text{ if } n > m$$

then

$$f(x) + g(x) = (a_0 + b_0) + (a_1 + b_1)x + \ldots + (a_m + b_m)x^m + \ldots + (a_n + b_n)x^n$$

$$a_{m+1} = a_{m+2} = a_n = 0$$

and if c ε K then

$$cf(x) = ca_0 + ca_1 x + ca_2 x^2 + \ldots. + ca_m x^m.$$

With these definitions of addition and scalar multiplication, it is easily proved as in Example 2.1.2 that K{x} is a vector space over K.

The zero vector in K{x} is the zero function 0(x) = 0. Also if f(x) ε K{x} then -f(x) is the additive inverse of f(x).

2.2 Subspaces

Definition 2.2.1

Let V be a vector space over a field K and W a subset of V. W is called a subspace of V if W is itself a vector space with respect to the vector addition and scalar multiplication defined on V. For practical purposes, when we need to check if a subset of a vector space is a subspace, we do not need to check all the axioms of a vector space on the given subset. The following theorem explains what is to be done.

Theorem 2.2.1

Let V be a vector space over a field K and W a subset of V.
W is a subspace of V if and only if

(i) $W \neq \varphi$.
(ii) if u, v ε W then u + v ε W
(iii) if u ε W and c ε K then cu ε W.

Proof:

Suppose (i), (ii) and (iii) in the theorem hold then all other axioms in Definition 2.1.1 of a vector space hold in W except axioms (ii) and (iii) since the vectors in W belong to V. We only need to show that there exists 0 ε W such that

$$u + 0 = u \text{ for all } u \in W$$

and for any u ε W there exists -u ε W such that

$$u + (-u) = 0.$$

Since $W \neq \varphi$ then there exists u ε W and since K is a field, 0 ε K and by condition (iii) of this theorem

$$u\,0 = 0 \in W \text{ and } u + 0 = u \text{ for all } u \in W.$$

Hence w satisfies condition (ii) of Definition 2.1.1. If v ε W and since K is a field, -1 ε K and by (iii) of this theorem

$$(-1)v = -v \in W \text{ and } v + (-v) = 0,$$

hence W satisfies condition (iii) of Definition 2.1.1. Thus W is a subspace of V. Conversely suppose W is a subspace of V then all the axioms (i), (ii) and (iii) in the theorem are clearly satisfied.

Remark:

In actual practice, to show that a subset of a vector space is a subspace, conditions (ii) and (iii) can be combined. Thus W is a subspace if

(i) $W \neq \varphi$ and
(ii) if $a, b \in K$ and $u, v \in W$ then $a\,u + b\,v \in W$.

Example 2.2.1

Consider $R^4 = \{(a_1, a_2, a_3, a_4) | a_i \in R\}$, by Example 2.1.2, R^4 is a vector space over R. If $W = \{(a_1, a_2, a_3, 0) | a_i \in R\}$ i.e. W is the set of all vectors in R^4 whose last coordinate is zero, show that W is a subspace of R^4.

Solution

Since the last coordinate of the vector $0 = (0, 0, 0, 0)$ is zero then

$$0 = (0, 0, 0, 0) \in W$$

and so

$$W \neq \varphi.$$

Let $u, v \in W$ where $u = (u_1, u_2, u_3, 0)$ and $v = (v_1, v_2, v_3, 0)$ then

$$u + v = (u_1 + v_1, u_2 + v_2, u_3 + v_3, 0 + 0) = (u_1 + v_1, u_2 + v_2, u_3 + v_3, 0)$$

Since the last coordinate of $u + v$ is zero then

$$u + v \in W.$$

If $u \in W$ where $u = (u_1, u_2, u_3, 0)$ and $c \in R$, then

$$cu = (cu_1, cu_2, cu_3, c0) = (cu_1, cu_2, cu_3, 0)$$

hence the last coordinate of cu is zero and so $cu \in W$. This shows that W is a subspace of R^4.

Example 2.2.2

Consider R^3 as a vector space over R. Let W be a subset R^3 defined by $W = \{(a_1, a_2, a_3) | a_1 = a_2 = a_3\}$. Show that W is a subspace of R^3.

Solution

Consider the zero vector in $R^3, 0 = (0, 0, 0)$.

In this vector the first coordinate is 0, second coordinate is 0 and the third co-ordinate is 0 hence in this case

$$a_1 = 0, a_2 = 0 \text{ and } a_3 = 0$$

and so $0 \in W$. If $u, v \in W$ then

$$u = (u_1, u_2, u_3) \text{ where } u_1 = u_2 = u_3$$

and

$$v=(v_1, v_2, v_3) \text{ where } v_1=v_2=v_3.$$

However

$$u+v=(u_1, u_2, u_3)+(v_1, v_2, v_3)=(u_1+v_1, u_2+v_2, u_3+v_3)$$

Since $u_1=u_2=u_3$ and $v_1=v_2=v_3$ then

$$u_1+v_1=u_2+v_2=u_3+v_3$$

hence $u+v \varepsilon W$. If $c \varepsilon R$ and $u \varepsilon W$ where $u=(u_1, u_2, u_3)$ then

$$cu=c(u_1, u_2, u_3)=(cu_1, cu_2, cu_3).$$

Since $u_1=u_2=u_3$ then

$$cu_1=cu_2=cu_3$$

hence,

$$cu \varepsilon W,$$

and so W is a subspace of V.

Let V a vector space over a field K and $v_1, v_2, \ldots, v_n$ vectors in V. If $a_1, a_2, \ldots, a_n$ are scalars in K then the vector $a_1v_1+a_2v_2+\ldots+a_nv_n$ is called a linear combination of the vectors $(v_1, v_2, \ldots, v_n)$.

Notice that for a vector $u \varepsilon V$ to be a linear combination of the vectors $v_1, v_2, \ldots, v_n$ it must be the sum of the scalar multiples of $v_1, v_2, \ldots, v_n$.
In other words there must exist $a_1, a_2, \ldots, a_n \varepsilon K$ such that $u=a_1v_1+a_2v_2+\ldots+a_nv_n$.

Example 2.2.3
Let u = (2, -1, -1), v = (1, -3, 2) and w = (1, 1, 2), write the vector y = (1, 20, -13) as a linear combination of vectors u, v and w.

Solution

To write y as a linear combination of u and v, we need to find scalars a, b and c such that y = au + bv + cw i.e. we are looking for real numbers a, b and c such that

$$\begin{aligned}(1, 20, -13) &= a(2, -1, -1)+b(1, -3, 2)+c(1, 1, 2)\\ &= (2a, -a, -a)+(b, -3b, 2b)+(c, c, 2c)\\ &= (2a+b+c, -a-3b+c, -a+2b+2c)\end{aligned}$$

hence

$$+1=2a+b+c$$

$$20=-a-3b+c$$

$$-13=-a+2b+2c.$$

On solving these three equations simultaneously we get

$$a=3, b=-7 \text{ and } c=+2$$

hence we write

$$(1, 20, -13)=3(2, -1, -1)-7(1, -3, 2)+2(1, 1, 2)$$

Theorem 2.2.2

Let V be a vector space over a field K and $v_1, v_2, \dots, v_n$ vectors in V. The set of all linear combinations of $v_1, v_2, \dots, v_n$ is a subspace of V.

Proof:

If W is the set of all linear combinations of the vectors $v_1, v_2, \dots, v_n$, we need to show that W is a subspace of V.

Since $0 = 0v_1 + 0v_2 + \dots + 0v_n$, then it is seen that 0 is a linear combination of $v_1, v_2, \dots, v_n$ hence

$$0 \varepsilon W$$

and so

$$W \underline{=} \varphi.$$

If u, v ε W then

$$u = a_1v_1 + a_2v_2 + \dots + a_n v_n$$

and

$$v = b_1v_1 + b_2v_2 + \dots + b_nv_n$$

where $a_i, b_i \varepsilon K$ then

$$\begin{aligned} u + v &= a_1v_1 + a_2v_2 + \dots + a_nv_n + b_1v_1 + b_2v_2 + \dots + b_nv_n \\ &= (a_1 + b_1)v_1 + (a_2 + b_2)v_2 + \dots + (a_n + b_n)v_n \end{aligned}$$

Since $a_i + b_i \varepsilon K$ then we have expressed u + v as a linear combination of $v_1, v_2, \dots, v_n$, hence

$$u + v \varepsilon W.$$

Lastly, if u ε W and c ε K where $u = a_1v_1 + a_2v_2 + \dots + a_nv_n$, then

$$cu = c(a_1v_1 + a_2v_2 + \dots + a_nv_n) = ca_1v_1 + ca_2v_2 + \dots + ca_nv_n$$

since c ε K and $a_i \varepsilon K$ then $ca_i \varepsilon K$ then

$$cu = ca_1v_1 + ca_2v_2 + \dots + ca_nv_n$$

is a linear combination of $v_1, v_2, \dots, v_n$ hence,

$$cu \varepsilon W.$$

W is therefore a subspace of V.

2.3 Basis and Dimension

Definition 2.3.1

Let V be a vector space over a field K and $v_1, v_2, \dots, V_n$ vectors in V. The vectors $v_1, v_2, \dots, v_n$ are linearly dependent over K if there exists scalars $a_1, a_2, \dots, a_n \varepsilon K$ not all of them equal 0 such that $a_1v_1 + a_2v_2 + \dots + a_nv_n = 0$.

Otherwise, the vectors are said to be linearly independent over K. Notice that the relation $a_1v_1 + a_2v_2 + \dots + a_nv_n = 0$ will always hold if the a_is are all 0.

However if the relation $a_1v_1+a_2v_2+...+a_nv_n=0$ hold only if $a_1=0, a_2=0, ..., a_n=0$ then we say that the vectors $v_1 v_2, ..., v_n$ are linearly independent.

Example 2.3.1

The vectors u = (2, -2, 0), v = (1, 3, 1) and w = (3, 1, 1) are linearly dependent because

$$u+v-w=(2,-2,0)+(1,3,1)-(3,1,1)=(0,0,0).$$

Example 2.3.2

Consider the vectors u = (3, 1, -4), v = (2, 5, 6) and w = (1, 4, 3), show that u, v, w are linearly independent.

Solution

To show that u, v, w are linearly independent, suppose for scalars a, b, c we have

$$au+bv+cw=0$$

then we need to show that a = 0, b = 0 and c = 0.

$$au+bv+cw=a(3,1,-4)+b(2,5,6)+c(1,4,3)=(0,0,0)$$

then

$$(3a+2b+c,\ a+5b+4c,\ -4a+6b+3c)=(0,0,0).$$

hence

$$3a+2b+c=0 \qquad (1)$$

$$a+5b+4c=0 \qquad (2)$$

$$-4a+6b+3c=0 \qquad (3)$$

multiply equation (2) by 4 and add to equation (3) to have

$$26b+19c=0 \qquad (4).$$

multiply equation (2) by -3 and add to equation (1) to have

$$-13b-11c=0 \qquad (5)$$

multiply equation (5) by 2 and add to equation (5) to have

$$-3c=0 \text{ hence } c=0.$$

Substitute c = 0 in equation (4) to have 26b = 0 hence

$$b=0.$$

We, lastly, substitute b = 0, c = 0 into equation (1) to have

$$3a+2.0+0,=0$$

hence

$$3a=0,$$

and so

$$a=0.$$

Since a = 0, b= 0 and c = 0 then the vectors u, v, and w are linearly independent.

Let V be a vector space over a field K. If $v_1, v_2, ..., v_n$ are vectors in V such that a vector u ε V can be written as a linear combination of $v_1, v_2, ..., v_n$, that is,

$$u=a_1v_1+a_2v_2+...+a_nv_n$$

then the scalars $a_1, a_2, a_3, ..., a_n$ are called coordinate of u with respect to $v_1, v_2, ..., v_n$.

Example 2.3.3

Find the coordinate of the vector (1, 0) with respect to the vectors (1, 1) and (-1, 2).

Solution

We need to find the scalars a, b such that

$$(1, 0) = a(1, 1) + b(-1, 2) = (a - b, a + 2b)$$

hence

$$a - b = 1$$
$$a + 2b = 0$$

subtracting first equation from the second we have

$$3b = -1,$$

hence

$$b = -\frac{1}{3}$$

substituting for b in the first equation, we have

$$a = 1 - \frac{1}{3} = \frac{2}{3}$$

hence the coordinates of (1, 0) with respect to (1, 1) and (-1, 2) is $\left(\frac{2}{3}, \frac{1}{3}\right)$.

Theorem 2.3.1

Let V be a vector space over a field K and $v_1, v_2, \ldots, v_n$ are nonzero vectors. The vectors are linearly dependent if and only if one of the vectors v_i can be written as a linear combination of the preceding vectors.

Proof:

Suppose

$$v_i = a_1v_1 + a_2v_2 + \ldots + a_{i-1}v_{i-1}$$

then subtracting v_i from both sides, we have

$$a_1v_1 + a_2v_2 + \ldots + a_{i-1}v_{i-1} - v_i = 0$$

hence

$$a_1v_1 + a_2v_2 + \ldots + a_{i-1}v_{i-1} - v_i + 0v_{i+1} + \ldots + 0v_n = 0$$

and the coefficient of v_i is -1 and it is not 0. This shows that the v_i's are linearly dependent.

Conversely suppose $v_1, v_2, \ldots, v_n$ are linearly dependent then there exists scalars $a_1, a_2, \ldots, a_n$ not all 0 such that

$$a_1v_1 + a_2v_2 + \ldots + a_nv_n = 0.$$

Let r be the largest integer such that $a_r \neq 0$ then we have

$$a_1v_1 + a_2v_2 + \ldots + a_rv_r + 0v_{r+1} + \ldots + 0v_n = 0$$

or

$$a_1v_1 + a_2v_2 + \ldots + a_rv_r = 0$$

hence

$$v_r = a_r^{-1}a_1v_1 + a_r^{-1}a_2v_2 + \ldots + a_r^{-1}a_{r-1}v_{r-1}$$

hence v_r has been written as a linear combination of the preceding vectors.

Definition 2.3.2

Let V be a vector space over a field K, the vectors $v_1, v_2, \ldots, v_n \, \varepsilon \, V$ are called basis of V if they are linearly independent and if every vector v ε V can be written as a linear combination of $v_1, v_2, \ldots, v_n$.

If every vector u ε V can be written as a linear combination of $v_1, v_2, \ldots, v_n$ then we say that $v_1, v_2, \ldots, v_n$ generate (or span) V. It therefore follows that the vectors $v_1, v_2, \ldots, v_n$ form a basis of the vector space V if $v_1, v_2, \ldots, v_n$ are linearly independent and they generate V.

Example 2.3.4

Consider vector space R^3 over R. Show that the vectors u = (1, 0, 0), v = (1, 1, 0) and w = (0, 1, 1) form a basis for R^3.

Solution

We need to show that the vectors u, v and w are linearly independent and that they generate R^3. To show they are linearly independent let

$$au + bv + cw = 0$$

then we need to show that a = 0, b = 0 and c = 0. We know that

$$au + bv + cw = 0$$

implies

$$a(1, 0, 0) + b(1, 1, 0) + c(0, 1, 1) = (0, 0, 0)$$
$$(a + b, b + c, c) = (0, 0, 0)$$

then

$$a + b = 0$$
$$b + c = 0$$
$$c = 0.$$

These three equations imply c = 0, b = 0 and a = 0 hence u, v and w are linearly independent. To show that they generate R^3, we pick an arbitrary vector q ε R^3 say q = (a, b, c) then we look for scalars x, y, z in R such that

$$(a, b, c) = xu + yv + zw = x(1, 0, 0) + y(1, 1, 0) + z(0, 1, 1)$$
$$(a, b, c) = (x + y, y + z, z)$$

The above equation gives

$$a = x + y$$
$$b = y + z$$
$$c = z$$

since $c = z$ then

$$\therefore\ y = b - z = b - c$$

and so

$$x = a - y = a - (b - c) = a - b + c,$$

hence

$$x = a - b + c,\ y = b - c \text{ and } z = c$$

we therefore have

$$(a, b, c) = (a - b + c)(1, 0, 0) + (b - c)(1, 1, 0) + c(0, 1, 1)$$

and so $(a, b, c) = q$ has been written as a linear combination of u, v, w. Since q is an arbitrary vector in V then it follows that any vector in R^3 can be so written hence u, v, w generate R^3. Since u, v, w are linearly independent and they generate R^3 then they form a basis for R^3.

Example 2.3.5

Consider vector space R^3 over R. Show that the vectors $u = (1, 1, 1)$, $v = (1, 2, 3)$ and $w = (2, -1, 1)$ form a basis for R^3.

Solution

We need to show that the three vectors are linearly independent and they generate R^3. To show they are linearly independent, let

$$au + bv + cw = 0,$$

then

$$a(1, 1, 1) + b(1, 2, 3) + c(2, -1, 1) = (0,0,0)$$
$$(a + b + 2c, a + 2b - c, a + 3b + c) = (0, 0, 0)$$

This implies that

$$a + b + 2c = 0 \quad (1)$$
$$a + 2b - c = 0 \quad (2)$$
$$a + 3b + c = 0 \quad (3)$$

If we subtract equation (1) from (2) and subtract equation (2) from (3) we have

$$b - 3c = 0 \quad (4)$$
$$b + 2c = 0 \quad (5)$$

If we subtract equation (4) from equation (5) we have

$$5c = 0$$

hence

$$c = 0.$$

If we substitute $c = 0$ into equation (5) we have

$$b + 2.0 = 0$$

hence

$$b = 0,$$

and lastly when we substitute $b = 0$ and $c = 0$ into equation (1) we have

$$a + 0 + 2.0 = 0$$

which implies

$$a = 0.$$

Then $a = 0$, $b = 0$ and $c = 0$ implies the vectors $(1, 1, 1)$, $(1, 2, 3)$ and $(2, -1, 1)$ are linearly independent. To show that these three vectors generate V, let $u \in R^3$ be any vector say, $u = (a, b, c)$.

We want scalars x, y, z such that

$$(a, b, c) = x(1, 1, 1) + y(1, 2, 3) + z(2, -1, 1) = (x + y + 2z, x + 2y - z, x + 3y + z)$$

hence

$$a = x + y + 2z \tag{1}$$
$$b = x + 2y - z \tag{2}$$
$$c = x + 3y + z \tag{3}$$

we then solve these three equations to find the values of x, y, z in terms of a, b and c.

Subtracting equation (1) from (2) and (2) from (3) we have

$$b - a = y - 3z \tag{4}$$
$$c - b = y + 2z; \tag{5}$$

Subtracting equation (4) from (5) we have

$$(c - b) - (b - a) = 5z.$$
$$c - 2b + a = 5z$$

hence

$$z = \frac{c - 2b + a}{5}$$

Substituting this into equation 5, we have

$$y = c - b - 2z = c - b - 2\left(\frac{c - 2b + a}{5}\right)$$

$$= \frac{5(c - b) - 2(c - 2b + a)}{5} = \frac{3c - b - 2a}{5}$$

Substituting the values of y and z in equation (1) we have

$$x = a - y - 2z$$

$$x = a - \frac{(3c - b - 2a)}{5} - \frac{2(a + 2b + a}{5}$$

$$= 5a - \frac{(3c - b - 2a) - 2(c + 2b + a}{5} = \frac{5a - 3b - 5c}{5}$$

hence

$$(a,b,c) = \frac{5a - 3b - 5c}{5}(1, 1, 1) + \frac{3c - b - 2a}{5}(1, 2, 3) + \left(\frac{a - 2b + c}{5}\right)(2, -1, 1).$$

This shows that the three vectors generate V, and so they form a basis for V. Examples 2.3.4 and 2.3.5 show that a vector space may have more than one basis. However the next theorem shows that when a vector space has two or more basis, the basis have the same number of elements.

Definition 2.3.3

If a vector space has finite number of elements in its basis then we say that the vector space is finite dimensional. Otherwise it is called infinite dimensional.

Theorem 2.3.2
Let V be a finite dimensional vector space. Then every basis of V has the same number of vectors.

Proof:
Let $(u_1, u_2, \ldots, u_n)$ be a basis of V and $(v_1, v_2, \ldots, v_m)$ be another basis of V we need to show that $n = m$.
Since $(u_1, u_2, \ldots, u_n)$ form a basis for V and $v_1 \varepsilon V$, then v_1 can be written as a linear combination of $(u_1, u_2, \ldots, u_n)$ hence the set $\{v_1, u_1, u_2, \ldots, u_n\}$ generate V and is linearly dependent. By theorem 2.2.2 one of the vectors in $\{v_1, u_1, u_2, \ldots, u_n]\}$ must be a linear combination of the preceding vectors. This vector cannot be v_1 because v_1 belongs to a set of linear independent vectors $\{v_1, v_2, \ldots, v_m\}$.

It follows that some vector $u_i \varepsilon \{v_1, u_1, u_2, \ldots, u_n\}$ is dependent on its predecessors, we delete this vector u_i from this set of vectors to have $\{v_1, u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n\}$ and this set still generate V. We repeat this argument with vector v_2 thus, since $\{v_1, u_1, \ldots u_{i-1}, u_{i+1}, \ldots, u_n\}$ generate V the set $\{v_1, v_2, u_1, \ldots, u_{i-1} u_{i+1}, \ldots, u_n\}$ is linearly dependent and also generate V.

Again by theorem 2.2.2, one of the vectors in $\{v_1, v_2, u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n\}$ is linear combination of the preceding vectors. This vector cannot be v_1 or v_2 because $\{v_1, v_2, \ldots, v_n\}$ is linearly independent hence it must be one of the u_i's. say u_k. Thus we can delete u_k to obtain

$$\{v_1, v_2, u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_{k-1}, u_{k+1}, \ldots, u_n\}.$$

We continue this argument with v_3 and so on. Each time, we add one of the vectors v_i's and delete one of the u_j's. If $m \leq n$ then finally we will obtain a generating set

$$\{v_1, v_2, \ldots, v_m, u_{i_1}, \ldots, u_{i_{n-m}}\}$$

It is not possible for $m > n$ because if $m > n$ then after n steps we obtain the generating set $\{v_1, v_2, \ldots, v_n\}$ and this will imply that the vector v_{n+1} will be a linear combination of $v_1, v_2, \ldots, v_n$ which is impossible since the v_i's are linearly independent. So $m \leq n$. Similarly if we take $(v_1, v_2, \ldots, v_m)$ as the basis and step by step introduce the vectors u_1 then u_2 and so on we obtain $n \leq m$. This will imply $n = m$.
The preceding theorem shows us that if a vector space has two bases then the number of elements in each basis are the same. This leads us to the following definition.

Definition 2.3.4
The dimension of a vector space is the number of elements in any of its basis.
This definition is independent of the basis of the vector space chosen because by the preceding theorem, all the basis of a vector space have the same number of elements.

Example 2.3.6

Let K be a field, by example 2.1.2, we showed that K^n is a vector space over K. The vectors

$$e_1 = (1, 0, 0,...,0,0)$$
$$e_2 = (0,1,0,...,0,0\}$$
$$e_n = (0,0,0,...,0,1)$$

form a basis of K^n. This basis is called the usual basis of K^n. Since there are n vectors in this basis of K^n it follows that the dimension of K^n is n.

Definition 2.3.5

Let V be a vector space over a field K and S a subset of V.
Let $v_1, v_2,..., v_m$ be vectors in S, we say that $\{v_1,..., v_m\}$ is a maximal linearly independent subset of S if (i) it is an independent subset of S and (ii) for any vector u ε S, $\{u, v_1, v_2,..., v_m\}$ is dependent.

This definition simply says that $\{v_1, v_2,..., v_n\}$ is a maximal linearly independent subset of S if they are independent and if we add one or more vectors of S to it, it becomes dependent.

Theorem 2.3.3

Let $\{v_1, v_2,..., v_n\}$ be a maximal linearly independent subset of V then $\{v_1, v_2,,..., v_n\}$ is a basis of V.

Proof:

Since $\{v_1, v_2,..., v_n\}$ is a maximal linearly independent subset of V then we only need to show that these vectors generate V since they are already linearly independent.
Let u ε V be arbitrary vector in V then $\{v_1, v_2,..., v_n, u\}$ is no more linearly independent hence by theorem 2.2.2, u can be written as a linear combination of the preceding vectors $(v_1, v_2,..., v_n\}$.

Since u is an arbitrary vector of V it follows that $\{v_1, v_2,..., v_n\}$ generate V hence it forms a basis of V.

2.3.4:

Let V be a vector space over a field K and W a subspace of V. If the dimension of V is n, then dim W ≤ n. If dim W = n then W = V.

Proof:

Since V has dimension n then the number of elements in any of its basis is n, any n + 1 or more vectors are linearly dependent. Notice that a basis of W must consist of linearly independent vectors and it cannot contain more than n vectors hence dim W ≤ n. If Dim W = n then W has n vectors $\{w_1, w_2,..., w_n\}$ in its basis.

Since this set is an independent set with n elements then it is also a basis of V hence W=V.

2.4 Sums and Direct Sums

Let V be a vector space over a field K, let U and W be subspaces of V. We define the sum of U and W as the set of all elements of the form u + w when u ε U and w ε W. We denote this sum by U +W, that is,

$$U + W = \{u + w \mid u \varepsilon U, w \varepsilon W\}$$

Theorem 2.4.1

Let V be a vector space over a field K and U, W subspaces of V then U + W is a subspace of V.

Proof:

To show that U + W is a subspace of V, we need to show that

(i) $U + W \neq \varphi$

(ii) If u + w and u′ + w′ are elements in U + W then their addition (u + w) + (u′ + w′) is an element of U + W.

(iii) We also need to show in conclusion that if c ε K and u + w ε U + W then

$$c(u + w) \varepsilon U + W.$$

Now since U and W are subspaces of V then

$$0 \varepsilon U \text{ and } 0 \varepsilon W$$

hence

$$0 + 0 \varepsilon U + W$$

and so

$$U + W \neq \varphi.$$

This proves (i).

Suppose u + w and u′ + w′ are elements in U + W where u, u′ ε U and w, w′ ε W then

$$(u + w) + (u' + w') = u + u' + w + w'$$

but U is a subspace of V then u + u′ ε U also W is a subspace of V then

$$w + w' \varepsilon W$$

hence

$$u + u' + w + w' \varepsilon U + W.$$

This shows that

$$(u + w) + (u' + w') \varepsilon U + W$$

and so (ii) is proved. If c ε K and u + w ε U + W, then

$$c(u + w) = cu + cw.$$

Again since U is a subspace of V and u ε U while c ε K, then

$$cu \varepsilon U.$$

Similarly since w ε W and c ε K and W is a subspace of V, then

$$cw \varepsilon W.$$

Since $cu \varepsilon U$ and $cw \varepsilon W$ then it follows that

$$cu + cw \varepsilon U + W$$

and so (iii) is proved. This shows that $U + W$ is a subspace of V.

Definition 2.4.1

Let V be a vector space over a field K, and U, W subspaces of V. V is said to be the direct sum of its subspaces U and W if the followings are satisfied.

(i) $V = U + W$ and

(ii) $U \cap W = \{0\}$.

If V is the direct sum of subspaces U and W we write this simply as $V = U \oplus W$.

Theorem 2.4.2

Let V be a vector space over a field K, let U and W be subspace of V then if V is the direct sum of U and W i.e. $V = U \oplus W$ then every element $v \varepsilon V$ can be written uniquely as the sum of an element $u \varepsilon U$ and an element $w \varepsilon W$.

Proof :

If $V = U \oplus W$ then by the definition of direct sum it follows that

$$V = U + W \text{ and } U \cap W = \{0\}.$$

Then $V = U + W$ implies that every element $v \varepsilon V$ can be written as a sum of an element $u \varepsilon U$ and an element $w \varepsilon W$. We now only need to show that such a sum can be written in one and only one way. On the contrary suppose $v \varepsilon V$ can be written as the sum of elements of U and W in two ways i.e. if

$$v = u + w \text{ and } v = u' + w'$$

where $u, u' \varepsilon U$ and $w, w' \varepsilon W$ then

$$u + w = v = u' + w'.$$

This implies that

$$u - u' = w' - w,$$

but $u - u' \varepsilon U$ since U is a subspace of V and $w' - w \varepsilon W$ since W is subspace of V hence $u - u' \varepsilon U$ and also $u - u' = w' - w \varepsilon W$, therefore

$$u - u' \varepsilon U \cap W = (0)$$

hence

$$u - u' = 0.$$

This implies $u = u'$.

Similarly, since

$$' - w \varepsilon W \text{ and } w' - w = u - u' \varepsilon U$$

then

$$w' - w \varepsilon U \cap W = \{0\}$$

hence

$$w' - w = 0$$

and so

$$w' = w.$$

This shows that $v \varepsilon V$ can be written uniquely as the sum of an element in U and an element in W.

Example 2.4.1

Consider R^3 as a vector space over R. Let $U=\{(0, 0, z)|z\varepsilon R\}$ and $W=\{(x, y, z) \mid x = z\}$

Show that $R^3 = U \oplus W$.

Solution

We need to show that $R^3 = U + W$ and $U \cap W = \{0\}$, let $v \varepsilon R^3$ where $v = (a, b, c)$ we want to find an element $u \varepsilon U$ and $w \varepsilon W$ such that $v = u + w$. I
In other words we want $x, y, z \varepsilon R$ such that

$$v = (a, b, c) = (0, 0, z) + (x, y, x) = (x, y, z + x)$$

it follows that

$$a = x,\ b = y \text{ and } c = x + z.$$

This implies $z = c - x = c - a$, hence

$$v = (a, b, c) = (0, 0, c - a) + (a, b, a).$$

Lastly we need to show that $U \cap W = \{0\}$. If $v \varepsilon U \cap W$ then

$$v \varepsilon U \text{ and } v \varepsilon W$$

then

$$v = (0, 0, z)$$

for $z \varepsilon R$ and $v \varepsilon (x, y, x)$ since $v \varepsilon W$, $v = (0, 0, z) = (x, y, x)$ implies $x = 0, y = 0$ and $x = z$, hence

$z = 0$.

We then have $x = 0$, $y = 0$ and $z = 0$ hence $v = 0$ and so

$$U \cap W = \{0\}.$$

2.5 Exercise

1. Let F be a field and $V = \{(a, b) \mid \text{and } b \varepsilon F\}$. If addition and scalar multiplication are defined as follows $(a, b) + (c, d) = (a + c, b + d)$ and $k(a, b) = (ka, kb)$ where (a, b) and (c, d) ε V and $k \varepsilon F$. Show that V is a vector space over F.
2. If V is a vector space over a field K show the following
   (i) if $c \varepsilon K$ and $0 \varepsilon V$ then $c0 = 0$.
   (ii) If $c \varepsilon K$ and $u \varepsilon V$ then $(-c)u = -(cu)$
   (iii) For $0 \varepsilon K$ and any $u \varepsilon V$, $0u = 0$
   (iv) If $cu = 0$ for $c \varepsilon K$ and $u \varepsilon V$ then $c = 0$ or $u=0$.
3. Consider R^4 as a vector space over R. Let W be a subset of R^4 defined as follows $W = \{(a, b, c, d) \mid a + b + c + d= 0\}$. Show that W is a subspace of R^4
4. Let $v_1 = (3, 1, -4)$, $v_2 = (2, 5, 6)$ and $v_3 = (1, 4, 3)$, write $u = (-7, 2, -7)$ as a linear combination of v_1, v_2 and v_3.

5. For what value of x is the vector (1, x, -8) a linear combination of the vectors (2, -1, -5) and (3, 0, -2).

6. Show that the vectors u = (1, 2, -3), v = (2, -1, -1) and w = (3, 1, 2) are linearly independent in R^3.

7. If V is a vector space over a field K and $\{v_1, v_2, v_3\}$ is a basis for V. If $u_1 = v_1, u_2 = v_1 + v_2$ and $u_3 = v_1 + v_2 + v_3$. Show that $\{u_1, u_2, u_3\}$ is also a basis for R^3.

8. If $v_1 = (3, 1, -4)$, $v_2 = (2, 5, 6)$ and $v_3 = (1, 4, 3)$, show that $\{v_1, v_2, v_3\}$ form a basis for R^3

9. In R^3, let W be a subspace generated by (1, 0, 0) and U the subspace generated by (1, 1, 0) and (0, 1, 1), show that $R^3 = U \oplus W$.

10. If $u_1, u_2, \ldots, u_n$ are linearly independent vectors in a vector space V. If $a_i \neq 0$ for each i, show that $(a_1u_1, a_2u_2, a_3u_3, \ldots, a_nu_n)$ are also linearly independent.

11. If u, v ε R^n such that u and v are linearly dependent, show that u is a scalar multiple of v.

12. (a) Given vectors u_1, u_2, u_3, u_4 in R^4 where $u_1 = (2, 1, 1, 6)$, $u_2 = (1, -1, -4, 0)$, $u_3 = (2, -1, -5, 2)$ and $u_4 = (1, 1, 2, 4)$. Are these vectors linearly independent?

(b) Find a basis of the subspace of R^4 generated by the vectors u_1, u_2, u_3, u_4 in 12(a).

13. Let V be a vector space of all 2 × 2 matrices over a field K. Show that V has dimension 4 by showing a basis of V having four elements.

14. Show that the vectors $u_1 = (0, 0, 0, 2)$, $u_2 = (0, 0, 1, 1,)$, $u_3 = (1, 0, 0, 4)$ and $u_4 = (1, 1, 0, 0)$ form a basis for R^4. Find the co-ordinates of each of the usual basis vectors in the ordered basis $\{u_1, u_2, u_3, u_4\}$.

15. If W is a subspace of C^4 generated by $u_1 = (1+i, 1, -1)$ and u_2 (1, 0, i)

(i) show that u_1 and u_2 form a basis for W

(ii) show that $w_1 = (1, i, 1+i)$ and $w_2 = (1, 1, 0)$ also form a basis for W.

(iii) Find the co-ordinates of u_1 and u_2 in the ordered basis $\{w_1, w_2\}$.

16. Let w = 2 + 3i and z = 1 - 2i. Show that w and z generate the complex field C as a vector space over field of real numbers R.

CHAPTER THREE

MATRICES

3.1 Definition and Examples

Let K be a field. A rectangular array of the form

$$A = \begin{pmatrix} a_{11}, a_{12}, \ldots, a_{1n} \\ a_{21}, a_{22}, \ldots, a_{2n} \\ \ldots, \ldots, \ldots, \ldots \\ a_{m1}, a_{m2}, \ldots, a_{mn} \end{pmatrix}$$

where $a_{ij} \varepsilon K$ is called a matrix over K. Capital letters are mostly used to denote a matrix while small letters are used for the scalar entries. The above matrix can simply be written as

$$A = (a_{ij}), i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, n$$

In the matrix A written above, the m horizontal n-tuples,

$$(a_{11}, a_{12}, \ldots, a_{1n},,). (a_{21}, a_{22}, \ldots, a_{2n}) \ldots . (a_{m1}, a_{m2}, \ldots a_{mn})$$

are called the rows of the matrix A. Thus $(a_{11}, a_{12}, \ldots, a_{1n})$ represents the first row, $(a_{21}, a_{22}, \ldots, a_{2n})$ represents the second row and $(a_{m1}, a_{m2}, \ldots, a_{mn})$ represents the m^{th} row. Then vertical m-tuples

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ \cdot \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ \cdot \\ \cdot \\ \cdot \\ a_{m2} \end{pmatrix}, \ldots, \begin{pmatrix} a_{1n} \\ a_{2n} \\ \cdot \\ \cdot \\ \cdot \\ a_{mn} \end{pmatrix}$$

are called columns of the matrix A. Thus

$$\begin{pmatrix} a_{11} \\ a_{22} \\ \cdot \\ \cdot \\ \cdot \\ a_{m1} \end{pmatrix}, \text{ is the first column.}$$

$$\begin{pmatrix} a_{12} \\ a_{22} \\ \cdot \\ \cdot \\ \cdot \\ a_{m2} \end{pmatrix} \text{ is the second column.}$$

and

$$\begin{pmatrix} a_{1n} \\ a_{2n} \\ \cdot \\ \cdot \\ \cdot \\ a_{mm} \end{pmatrix} \text{ is the nth column.}$$

A matrix with m number of rows and n number of columns is called an m by n matrix simply written as m × n matrix.

The element a_{ij} called the ij-entry or ij-component appears in the ith row and jth column. There are many types of matrices.

(a) When the number of rows is 1 then the matrix $A = (a_1, a_2, \ldots, a_n)$ and is simply called a row matrix.

(b) If the number of columns is 1 then the matrix $A = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{pmatrix}$ is called a column matrix.

(c) An m × n matrix whose entries are all equal to zero is called the zero matrix or null matrix of order m × n.

(d) When the number of rows of a matrix is the same as the number of columns then we call the matrix a square matrix i.e. if m = n then we have an n × n matrix called a square matrix of order n or an n-square matrix.

(e) A diagonal matrix of order n in which all the entries outside the diagonal entries are all zero is called a diagonal matrix of order n. For instance

$$A = \begin{pmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & 0 \\ 0 & 0 & 0 & a_{44} \end{pmatrix}$$

A is a diagonal matrix of order 4.

(f) An identity matrix is a diagonal matrix where the diagonal entries are all equal to 1 and is denoted by I or I_n For instance

$$I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

is an identity matrix of order 4.

3.1 Equality of Matrices, Addition and Scalar Multiplication

Let $A = (a_{ij})$ and $B = (b_{ij})$ be matrices over a field K, the two matrices are said to be equal if they have the same number of rows and same number of columns and if all the corresponding entries are equal.

In other words, $A = (a_{ij}), B = (b_{ij})$ if they have the same order and $a_{ij} = b_{ij}$ for all i,j. We will simply write.

$$A = B.$$

Let $A = (a_{ij})$ and $B = (b_{ij})$ be matrices, these two matrices can be added only when they have the same order. We define the addition of A and B as the matrix obtained by adding the corresponding entries. So

$$A + B = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$$

that is, if

$$A = \begin{pmatrix} a_{11}\, a_{12} \cdots a_{1n} \\ a_{21}\, a_{22} \cdots a_{2n} \\ \cdot \\ \cdot \\ \cdot \\ a_{m1}\, a_{m2} \cdots a_{mn} \end{pmatrix} \text{ and } B = \begin{pmatrix} b_{11} & b_{12} & \cdot\ \cdot\ \cdot & b_{1n} \\ b_{21} & b_{22} & \cdot\ \cdot\ \cdot & b_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ b_{m1} & b_{m2} & \cdot\ \cdot\ \cdot & b_{mn} \end{pmatrix}$$

then

$$A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_m1 + b_m1 & a_m2 + b_m2 & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

Example 3.1.1

Let $A = \begin{pmatrix} 3 & 1 & 2 \\ 5 & 7 & 9 \end{pmatrix}$ and $B = \begin{pmatrix} 4 & 0 & 3 \\ 2 & 5 & -2 \end{pmatrix}$ then

$$A+B=\begin{pmatrix}3+4 & 1+0 & 2+3\\ 5+2 & 7+5 & 9-2\end{pmatrix}=\begin{pmatrix}7 & 1 & 5\\ 7 & 12 & 7\end{pmatrix}$$

Let c be a scalar and A = (a_{ij}) an m × n matrix, then the product of A by a scalar c denoted by cA is the matrix obtained by multiplying each entry of A by c. That is,

$$cA=(ca_{ij})=\begin{pmatrix}ca_{11} & ca_{12} & \cdots & ca_{1n}\\ ca_{21} & ca_{22} & \cdots & ca_{2n}\\ \vdots & & & \vdots\\ \vdots & & & \vdots\\ ca_{m1} & ca_{m2} & \cdots & ca_{mn}\end{pmatrix}$$

Example 3.1.2

If $A=\begin{pmatrix}7 & 3 & 2\\ 4 & -1 & 5\\ 2 & 1 & 8\end{pmatrix}$ then

$$3A=\begin{pmatrix}3\times7 & 3\times3 & 3\times2\\ 3\times4 & 3\times(-1) & 3\times5\\ 3\times2 & 3\times1 & 3\times8\end{pmatrix}=\begin{pmatrix}21 & 9 & 6\\ 12 & -3 & 15\\ 6 & 3 & 24\end{pmatrix}$$

Observe that if A and B are m × n matrices then A + B is also an m × n matrix and if c is a scalar then cA is also an m × n matrix. We define

$$-A=-1.A$$

and

$$A-B=A+(-B)$$

Thus -A is defined as a matrix whose entries are the negative of the corresponding entries of A. If A = (a_{ij}) then $-A=(-a_{ij})$ and if B = b_{ij} then

$$A-B=(a_{ij})+(-b_{ij})=(a_{ij}-b_{ij})$$

Example 3.1.3

Let A = $\begin{pmatrix}1 & 3 & 0\\ -4 & 2 & 5\end{pmatrix}$, B = $\begin{pmatrix}1 & 1 & -2\\ 0 & 1 & -1\end{pmatrix}$ and C = $\begin{pmatrix}5 & 1 & 2\\ 3 & -1 & 1\end{pmatrix}$ then

$$2A-3B+4C=\begin{pmatrix}2 & 6 & 0\\ -8 & 4 & 10\end{pmatrix}-\begin{pmatrix}3 & 3 & -6\\ 0 & 3 & -3\end{pmatrix}+\begin{pmatrix}20 & 4 & 8\\ 12 & -4 & 4\end{pmatrix}=\begin{pmatrix}19 & 7 & 14\\ 4 & 3 & 17\end{pmatrix}$$

3.3 Multiplication of Matrices

Let A and B be two matrices, matrix A can be multiplied by matrix B in that order to have a matrix AB if the number of columns of A is equal to the number of rows of B. Suppose A is an m × n matrix then for the product AB to exist, the matrix B must have an order of the form n × r where r is an integer.

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \text{ and } B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ b_{n1} & b_{n2} & \cdots & b_{nr} \end{pmatrix}$$

the product AB is an m × r matrix obtained by taking the scalar product of the row vector $(a_{i1}, a_{i2}, \ldots, a_{in})$ from matrix A and the column vector

$$\begin{pmatrix} b_{1j} \\ b_{2j} \\ \cdot \\ \cdot \\ \cdot \\ b_{nj} \end{pmatrix}$$

to give us

$$a_{i1} b_{1j} + a_{i2} b_{2j} + \ldots + a_{in} b_{nj} = \sum_{t=1}^{n} a_{it} b_{tj}$$

It follows that if $AB = C = (c_{ij})$, then the ij-entry

$$(c_{ij}) = \sum_{t=1}^{n} a_{it} b_{tj} .$$

This shows that to compute the ij-entry in the product AB, we look for the i^{th}, row in matrix A and j^{th} column in matrix B then we compute the scalar product. Note that if A is an m × p matrix and B is an s × r matrix, then the product AB is not defined if $p \neq s$.

Example 3.3.1

Let $A = \begin{pmatrix} a\,b \\ c\,d \\ e\,f \end{pmatrix}$ and $B = \begin{pmatrix} g\,h\,i \\ m\,n\,p \end{pmatrix}$ then

$$AB = \begin{pmatrix} a\,b \\ c\,d \\ e\,f \end{pmatrix} \begin{pmatrix} g\,h\,l \\ m\,n\,p \end{pmatrix} = \begin{pmatrix} ag + bm, ah + bn, al + bp \\ cg + dm, ch + dn, cl + dp. \\ eg + fm, eh + fn, el + fp \end{pmatrix}$$

Example 3.3.2

If $A = \begin{pmatrix} 2\,3 \\ 4\,5 \end{pmatrix}$ and $B = \begin{pmatrix} -1\,3\,4 \\ 2\,4\,6 \end{pmatrix}$ then

$$AB=\begin{pmatrix}a\,b\\c\,d\\e\,f\end{pmatrix}\begin{pmatrix}g\,h\,l\\m\,n\,p\end{pmatrix}=\begin{pmatrix}ag+bm,\ ah+bn,\ al+bp\\cg+dm,\ ch+dn,\ cl+dp\\eg+fm,\ eh+fn,\ el+fp\end{pmatrix}$$

If A is an m × n matrix and B is an n × p matrix, then the product AB is an m × p matrix. The product BA can only exist if $p=m$. This shows that in general, the product AB of two matrices may be defined while the product BA may not be defined. In fact even when the products AB and BA are defined they may not be equal.
If A is a p × r matrix while B is an r × p matrix where $p \neq r$ then AB is a p × p matrix while BA is an r × r matrix. In this case AB and BA cannot be equal because they do not even have the same order. Consider the following example.

Example 3.3.3

If $=\begin{pmatrix}3\,4\\5\,6\end{pmatrix}$ and $B=\begin{pmatrix}1\,2\\4\,5\end{pmatrix}$ then

$$AB=\begin{pmatrix}3\,4\\5\,6\end{pmatrix}\begin{pmatrix}1\,2\\4\,5\end{pmatrix}=\begin{pmatrix}19\,26\\29\,40\end{pmatrix},\quad BA=\begin{pmatrix}1\,2\\4\,5\end{pmatrix}\begin{pmatrix}3\,4\\5\,6\end{pmatrix}=\begin{pmatrix}13\,16\\37\,46\end{pmatrix}$$

Since $\begin{pmatrix}19\,26\\29\,40\end{pmatrix}\neq\begin{pmatrix}13\,16\\37\,46\end{pmatrix}$ then $AB \neq BA$.

The preceding remark and example show that multiplication of matrices is not commutative.

3.4 Transpose, Row Equivalence and Elementary Row Operations

Let A be an m × n matrix, the transpose of A (denoted by A^t) is the matrix obtained by writing the rows of A as columns and its columns as rows. This implies that A^t will be an n × m matrix. If

$$A=\begin{pmatrix}a_{11}\,a_{12}\cdots a_{1n}\\a_{21}\,a_{22}\cdots a_{2n}\\\cdot\qquad\cdot\\\cdot\qquad\cdot\\\cdot\qquad\cdot\\a_{m1}\,a_{m2}\cdots a_{mn}\end{pmatrix}$$

then,

$$A^{t}=\begin{pmatrix}a_{11} & a_{21} & \cdots & a_{m1}\\a_{12} & a_{22} & \cdots & a_{m2}\\\cdot & & & \cdot\\\cdot & & & \cdot\\\cdot & & & \cdot\\a_{1n} & a_{2n} & \cdots & a_{mn}\end{pmatrix}$$

Example 3.4.1

If $A = \begin{pmatrix} 3 & 4 \\ 5 & 6 \\ -8 & 0 \end{pmatrix}$ then $A^{t} = \begin{pmatrix} 3 & 5 & -8 \\ 4 & 6 & 0 \end{pmatrix}$.

Theorem 3.4.2

Let A and B be matrices then

(i) $(A+B)^{t} = A^{t} + B^{t}$

(ii) $(AB)^{t} = B^{t} A^{t}$.

Proof:

(i)Let $A = (a_{ij})$ and $B = (b_{ij})$ $i = 1, 2, 3, \ldots, m$, $j = 1, 2, 3, \ldots, n$. Notice that A and B can be added only when they have the same order so we assume that A and B are of order $m \times n$.

$$A + B = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$$

$$\begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{pmatrix}$$

$$(A+B)^{t} = \begin{pmatrix} a_{11}+b_{11} & a_{21}+b_{21} & \cdots & a_{m1}+b_{m1} \\ a_{12}+b_{12} & a_{22}+b_{22} & \cdots & a_{m2}+b_{m2} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a_{1}n+b_{12} & a_{2n}+b_{2n} & \cdots & a_{mn}+b_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}+b_{11} & a_{21}+b_{21} & \cdots & a_{m1}+b_{m1} \\ a_{12}+b_{12} & a_{22}+b_{22} & \cdots & a_{m2}+b_{m2} \\ & & & \cdot \\ & & & \cdot \\ a_{1n}+b_{1n} & a_{2n}+b_{2n} & \cdots & a_{m2}+b_{mn} \end{pmatrix}$$

hence,

$$(A+B)^t=A^t+B^t.$$

(ii) Let $A=(a_{ij})$ be an $n\times m$ matrix while $B=(b_{jk})$ be an $m\times r$.

The ij^{th}-entry, which is the element in the i^{th} row and j^{th} column of the matrix AB is

$$a_{i1}b_{1j}+a_{i1}b_{2j}+\ldots+a_{im}b_{mj}.$$

This is exactly the element which appears in the j^{th} row and i^{th} column of $(AB)^t$, that is

$$(AB)^t=(C_{ij})=\left(\sum_{k=1}^{m}a_{ik}b_{kj}\right).$$

However the jth row of B^t are the elements from the jth column of B, and it is $(b_{1j},b_{2j},\ldots,b_{mj})$ The i^{th} column of A^t consists of elements from the i^{th} row of A, and it is

$$\begin{pmatrix}a_{i1},\\ a_{i1}\\ a_{im}\end{pmatrix}$$

Hence the element in the j^{th} row and i^{th} column of the matrix B^tA^t is the scalar product of the two vectors.

$$\begin{pmatrix}a_{i1}\\ a_{i2}\\ \vdots\\ a_{im}\end{pmatrix}\text{ and }(b_{1j},b_{2j},\ldots,b_{mj})$$

which gives

$$a_{i1}b_{ij}+a_{i2}b_{2j}+\ldots+a_{im}b_{mj}=\sum_{k=1}^{m}a_{ik}b_{kj}=C_{ij}$$

It then follows that

$$(AB)^t=B^tA^t.$$

Definition 3.4.1

Let $A=(a_{ij})$ be an $m\times n$ matrix. A is called an echelon matrix or it is in echelon form if the number of zero entries before the first non-zero entry of a row increases row by row until only zero rows remain.

The first non-zero row entry in a row of an echelon matrix is called the distinguished element on that row.

Example 3.4.2
The following matrices are examples of echelon matrices.

$$A = \begin{pmatrix} 3 & 2 & 4 & 0 & 1 \\ 0 & 0 & 6 & -2 & 5 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 4 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

In matrix A the distinguished elements are 3, 6 and 1 while in matrix B, the distinguished elements are 1, 4 and 1.

Definition 3.4.2
Given a matrix A, each of the following operations when performed on any of the rows of matrix A, is called elementary row operation.

(a) Interchange of ith row and jth row, which we will simply denote by R_i implies R_j

(b) Multiply by non-zero scalar p the ith row, simply denoted by $R_i \Rightarrow pR_i$ where $p \neq 0$.

(c) Replace the ith row by the sum of ith row and the product of p and jth row.

$$\text{i.e. } R_i \Rightarrow R_i + pR_j.$$

Operations (b) and (c) can be combined into one single operation namely.

(d) Replace ith row by the sum of p times ith row and q times jth row.

$$\text{i.e. } R_i \Rightarrow pR_i + qR_j.$$

Definition 3.4.3
Let A and B be matrices, matrix A is said to be row equivalent to matrix B if B can be obtained from A by performing a finite sequence of elementary row operations on A. The following example shows that any matrix is row equivalent to an echelon matrix.

Example 3.4.3

If If $A = \begin{pmatrix} 3 & 1 & 2 & 1 \\ 2 & 2 & 1 & -2 \\ 1 & 3 & -1 & 3 \end{pmatrix}$ reduce A to echelon form. reduce A to echelon form.

Solution

To reduce A to echelon form we first perform elementary row operations on the second and third rows to reduce the entry 2 in the second row and first column to 0 and also the entry 1 in the third row and first column to 0. So, we apply the operations.

$$R_2 \Rightarrow -2R_3 + R_2 \text{ and } R_3 \rightarrow -3R_3 + R_1$$

to reduce A to echelon form. A to $\begin{pmatrix} 3 & 1 & 2 & 1 \\ 0 & -4 & 3 & -8 \\ 0 & -8 & 5 & -8 \end{pmatrix}$

This form is not yet in echelon form because we need to reduce the entry -8 in the third row, second column to zero. To achieve this, we can perform the follow elementary row operation on it: $R_3 \Rightarrow -2R_2 + R_3$.

$$\begin{pmatrix} 3 & 1 & 2 & 1 \\ 0 & -4 & 3 & -8 \\ 0 & -8 & 5 & -8 \end{pmatrix} \text{ to } \begin{pmatrix} 3 & 1 & 2 & 1 \\ 0 & -4 & 3 & -8 \\ 0 & 0 & -1 & 8 \end{pmatrix}$$

The echelon form is

$$\begin{pmatrix} 3 & 1 & 2 & 1 \\ 0 & -4 & 3 & -8 \\ 0 & 0 & -1 & 8 \end{pmatrix}$$

The next theorem will be frequently used later.

Theorem: 3.4.3

Let A be an echelon matrix with non-zero rows $R_1, R_2, \ldots, R_m$, then $R_1, R_2, \ldots, R_m$, are linearly independent.

Proof:

Suppose the non-zero rows $R_m, R_{m-1}, \ldots, R_2, R_1$ are linearly dependent then by theorem 1.3.1 one of the rows is a linear combination of the preceding rows. Let this row be R_s where $s < m$. It then follows that

$$R_s = a_{s+1}R_{s+1} + a_{s+2}R_{s+2} + \ldots + a_mR_m.$$

Since the matrix A is in echelon form, the first few entries of R_s are zeros, let the first non-zero entry (i.e. the distinguished entry) in R_s be in the nth component of R_s i.e. a_{sr} is the distinguished entry of the s-row R_s.

However, since the matrix is in echelon form, all the mth components of the rows $R_{s+1}, R_{s+2}, \ldots, R_m$ are all zeros hence the r^{th} component of R_s which is $a_{sr} = a_{s+1}.0 + a_{s+2}.0 + \ldots + a_n.0 = 0$.

This shows that $a_{sr} = 0$ and it contradicts the fact that $a_{sr} \underline{=} 0$ being the distinguished element in R_s. Thus $R_1, R_2, \ldots, R_m$ are linearly independent.

3.5 Applications

Reduction of a given matrix to echelon form and the fact that the non-zero rows of an echelon matrix are linearly independent, can be applied to solve some problems that occur in linear algebra. In Chapter One, we discussed a method of checking whether or not a given number of vectors in a vector space are linearly independent or not. The idea of echelon matrix can be used to check linear dependence or linear independence of such vectors in a vector space. The procedure is as follows: Let V be an n-dimensional vector space, we need to check if the given vectors $v_1, v_2, \ldots, v_n$ are linearly independent. We will form a matrix A by making each vector v_i the i^{th} row of that matrix A i.e. first row (R_1) of A is vector v_1, second row (R_2) of A is vector v_2 and so on. The last row (nth row) of A is the vector v_n. We then obtain an n × n matrix.

$$A = \begin{pmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ \cdot \\ v_n \end{pmatrix} \text{ where } \begin{array}{l} v_1 = a_{11}, a_{12}, \ldots, a_{1n} \\ v_2 = a_{21}, a_{22}, \ldots, a_{2n} \\ \vdots \\ v_n = a_{n1}, a_{n2}, \ldots, a_{nn} \end{array}$$

Next we row reduce A to an echelon form.

If the echelon matrix we obtain from A, has no zero row then all the rows i.e.$(v_1, v_2, \ldots, v_n)$ are linearly independent.

But if the echelon matrix has at least one zero row then the vectors are linearly dependent and by theorem 3.4.3, only the non-zero rows are linearly independent.

Example 3.5.1

Determine if the following vectors are linearly independent; $u_1 = (2,3,4)$, $u_2 = (3,0,2)$ and $u_3 = (3,1,4)$.

Solution

We form a matrix A whose rows are the given vectors, then

$$A = \begin{pmatrix} 2,3,4 \\ 3,0\,2 \\ 3\,1\,4 \end{pmatrix}$$

We then row reduce A to echelon form using the method earlier discussed.

R_2 implies $R_3 - R_2$ and R_3 implies $-3R_1 + 2R_3$ then

$$A \underset{\Rightarrow}{\overset{\text{to}}{\longrightarrow}} \begin{pmatrix} 2 & 3 & 4 \\ 0 & 1 & 2 \\ 0 & -7 & -4 \end{pmatrix}$$

This is not yet in echelon form, for we need to reduce -7 to zero, and this can be done as follows

$$R_3 \Rightarrow 7R_2 + R_3$$

then

$$\begin{pmatrix} 2 & 3 & 4 \\ 0 & 1 & 2 \\ 0 & -7 & -4 \end{pmatrix} \text{to} \begin{pmatrix} 2 & 3 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 10 \end{pmatrix}$$

The echelon form obtained from A is

$$\begin{pmatrix} 2 & 3 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 10 \end{pmatrix}$$

Since there are no zero rows in the reduced form, it follows that the rows of A are linearly independent, hence the vectors (2, 3, 4), (3, 0 2) and (3, 1, 4) are actually linearly independent.

This method is also useful in establishing the fact that some vectors are not independent.

Example 3.5.2

Consider vector $u_1 = (2, 5, -1, 2)$, $u_2 = (1, 2, 0, -1)$, $u_3 = (2, 6, -3, -3)$ and $u_4 = (3, 10, -6, -5)$. Check if the vectors v_1, v_2, v_3, v_4 are linearly independent or not.

Solution

We form a matrix A by placing each vector u_i as the ith row of A to have

$$A = \begin{pmatrix} 2 & 5 & -1 & 2 \\ 1 & 2 & 0 & -1 \\ 2 & 6 & -3 & -3 \\ 3 & 10 & -6 & -5 \end{pmatrix}$$

We then row reduce A to echelon from

$$R_2 \Rightarrow -2R_2 + R_1$$
$$R_3 \rightarrow 2R_2 + R_3$$
$$R_4 \rightarrow 3R_2 + R_4$$

then

$$A \text{ to } \begin{pmatrix} 2 & 5 & -1 & 2 \\ 0 & 1 & -1 & 4 \\ 0 & 2 & -3 & -1 \\ 0 & 4 & -6 & -2 \end{pmatrix}$$

This form is not yet echelon form and so no conclusion can be reached at this point. We further reduce the matrix as follows:

$$R_4 \overset{-}{\rightarrow} 2R_3 + R_4,$$

$R_3 = R_3 - 2R_2$

$R_4 = R_4 - 2R_3$

hence

$$\begin{pmatrix} 2 & 5 & -1 & 2 \\ 0 & 1 & -1 & 4 \\ 0 & 2 & -3 & -1 \\ 0 & 4 & -6 & -2 \end{pmatrix} \text{to} \begin{pmatrix} 2 & 5 & -1 & 2 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & -1 & -9 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$R_4 = -2R_3 + R_4$

The echelon form of the matrix is

$$\begin{pmatrix} 2 & 5 & -1 & 2 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & -1 & -9 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

This echelon form has a zero row (the last row is zero) hence the rows of A are dependent. This shows that the vectors (2, 5, -1, 2), (1, 2, 0, -1), (2, 6, -3, -3) and (3, 10, -6, -5) are linearly dependent.

Definition 3.5.1

Let $A = (a_{ij})$ be an $n \times n$ matrix if $R_1, R_2, \ldots, R_n$ are the n rows of A, the vector space V spanned by these row vectors is called the row space of the matrix A.

We need to find the basis for this vector space and to do this we only need to row reduce this matrix to echelon form. By theorem 3.4.3 the non-zero rows of A when reduced to echelon form are linearly independent. Since the rows, by definition, generate V, then the non-zero rows form a basis for the row space of A = V. The dimension of this row space of A is called the rank of A. This leads us to the following definition.

Definition 3.5.2

The rank of a matrix A is the number of non-zero rows of the matrix, when reduced to echelon form.

Given a vector space V and a subspace W of the vector space, if we are given the generators of this subspace we can apply the preceding method to determine the dimension and the basis for the subspace W.

Example 3.5.3

Consider R^4 as a vector space over R. Let W be a subspace of R^4 generated by the vectors, $u_1 = (-1, 2, -4, -1)$, $u_2 = (-2, 3, -9, 1)$, $u_3 = (2, 0, 12, -10)$ and $u_4 = (4, -10, 14, 10)$. Find a basis and dimension of W.

Solution

In the question, it is stated that the vectors u_1, u_2, u_3, u_4 generate W. The question did not tell us that they are linearly independent for if they are linearly independent and since they generate W, then they will form a basis for W and in this case, dimension will be 4. The first thing we do is to find out, how many of these four vectors are linearly independent. To do this we form a matrix A whose rows are the vectors u_1, u_2, u_3 and u_4 to have

$$A = \begin{pmatrix} -1 & 2 & -4 & -1 \\ -2 & 3 & -9 & 1 \\ 2 & 0 & 12 & -10 \\ 4 & -10 & 14 & 10 \end{pmatrix}$$

We then row reduce A to echelon form, using the following operations.

$$R_2 \to R_2 + R_3$$

$$R_3 \to 2R_1 + R_3$$

$$R_4 \to 2R_2 + R_4$$

and we get the matrix

$$\begin{pmatrix} -1 & 2 & -4 & -1 \\ 0 & 3 & 3 & -9 \\ 0 & 4 & 4 & -12 \\ 0 & -4 & -4 & 12 \end{pmatrix}$$

$-36 + (-36$

Furthermore, to reduce the above further, we can perform the following operations.

$$R_3 \to 4R_2 \cancel{+} 3R_3$$

$$R_4 \to R_3 + R_4$$

This gives

$$\begin{pmatrix} -1 & 2 & -4 & 1 \\ 0 & 3 & 3 & -9 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

This is the required echelon matrix. There are two zero rows and two non-zero rows. This implies that only two vectors are linearly independent among u_1, u_2, u_3, u_4 that were given initially. Since the dimension of a vector space is the number of its linearly independent vectors then dim(W) = 2 and its basis are (-1, 2, -4, 1) and (0, 3, 3, -9).

Example 3.5.4

Find the rank of the matrix $A = \begin{pmatrix} -3 & 0 & 3 & 2 \\ 2 & 1 & -1 & 4 \\ 1 & 2 & -2 & -1 \end{pmatrix}$.

Solution

We row-reduce to echelon form.

$$\begin{pmatrix} -3 & 0 & 3 & 2 \\ 2 & 1 & -1 & 4 \\ 1 & 2 & -2 & -1 \end{pmatrix} \text{ to } \begin{pmatrix} -3 & 0 & 3 & 2 \\ 0 & -3 & 3 & 16 \\ 0 & -3 & 3 & 6 \end{pmatrix} \text{ to } \begin{pmatrix} -3 & 0 & 3 & 2 \\ 0 & 3 & 3 & 16 \\ 0 & 0 & 6 & 22 \end{pmatrix}$$

Thus rank (A) = 3.

Definition 3.5.3

An $n \times n$ matrix A is called invertible if there exists an $n \times n$ matrix B. Such that $AB = I_n$. Matrix B is called the inverse of A and is denoted by A^{-1}. We shall discuss methods of computing inverse of an invertible matrix in Chapter Four.

3.6 Exercises

1. Let V be vector space of $n \times n$ matrices over a field K. If $U = \{B \varepsilon V | B = B^t\}$ and $W = \{A \varepsilon V | A^t = -A\}$. Show that U and W are subspaces of V.

2. Matrix A is called a symmetric matrix if $A^t = A$. Show that for any $n \times n$ matrix A, $A + A^t$ and AA^t are symmetric matrices.

3. Let V be a vector space of 2×2 symmetric matrices over the set of real numbers R.
(i) Show that the following matrices form a basis for V.

$$\left\{\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}\right\}$$

(ii) Find the coordinate vector of the matrix

$$\begin{pmatrix} 1 & -5 \\ -5 & 5 \end{pmatrix}$$

in the above basis.

4. A subspace W of R^5 is generated by the vectors. $u_1 = (-2, -7, -3, -6, -13)$, $u_2 = (-1, -3, 2, -5, -4)$, $u_3 = (1, 4, 1, 3, 5)$ and $u_4 = (1, 4, 2, 4, 3)$. Find the dimension and the basis of W.

5. Determine the rank of the matrix $A = \begin{pmatrix} 1 & 3 & -1 & 0 \\ 0 & 11 & 5 & 5 \\ 2 & -5 & 3 & 5 \\ 4 & 1 & 1 & 5 \end{pmatrix}$.

6. If U is a subspace of R^5 generated by $\{u_1 = (1, 3, -1, -5, -6)$, $u_2 = (1, 5, 2, -2, 3)$ nd $u_3 = (2, 3, -1, -5, -5)\}$ and W is also a subspace of V generated by $\{w_1 = (2, 9, 0, -5, -2)$, $w_2 = (1, 3, -3, -1, -4)$ and $w_3 = (1, 4, -1, -2, -2)\}$. Find the dimension and basis of U + W.

7. If $A = \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix}$, then compute A^n

8. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, find A if A commutes with the matrix $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

9. W is the solution space of the following homogeneous equations

$$x + 3y + 2z = 0$$
$$3x + 5y + 8z = 0$$
$$x + 5y + 8z = 0.$$

Find a basis and the dimension of W.

10. Let V be a space generated by the following polynomials.
(i) $x^3 - 4x^2 + 2x + 3$ (ii) $2x^3 - x^2 - 3x + 5$ (iii) $x^3 + 2x^2 + 4x - 1$
Find a basis and dimension of V.

11. Let $A = \begin{pmatrix} 1 & -2 & -1 \\ 3 & -4 & 5 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -1 & 3 \\ 2 & -1 & 10 \\ 3 & -5 & 1 \end{pmatrix}$ show that A and B have the same row space.

12. If $A = \begin{pmatrix} 1 & 3 \\ 5 & 3 \end{pmatrix}$, find the vector $u = \begin{pmatrix} x \\ y \end{pmatrix}$ such that $Au = 6u$.

13. Let V be a vector space of all $n \times n$ matrices over a field F. If A is a given element in V and $W = \{M \varepsilon V \mid AM = MA\}$ show that W is a subspace of V.

14. (i) Let V be the vector space of symmetric 2×2 matrices over R. Show that the matrices

$$\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}$$

form a basis for V.

(ii) If $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$, find coordinate vector of A relative to the above basis of V.

15. Let W be a subset of R^4 where $W = \{a, b, c, d) \mid a = d, b = 2c\}$

(i) show that W is a subspace

(ii) Find the basis and dimension of W.

CHAPTER FOUR

LINEAR MAPPINGS

4.1 MAPPING

Let A and B be any sets, a mapping f from set A to set B is an association which associates to every element of A a unique element of B. Mapping is also called function. We will denote this function simply by

$$f: A \longrightarrow B.$$

If f: $A \longrightarrow B$ is a mapping and a ε A then we will denote the unique element in B to which a ε A is associated by f(a). We can also call f(a) the value of f at a or the image of a ε A under f. A is called the domain of f and B the co-domain. The set of all elements f(a) for all a ε A is called the image of A and is denoted by f(A), hence $f(A)=\{f(a)|a\varepsilon A\}$.

Example 4.1.1

Consider the set R of real numbers and a map f: $R \longrightarrow R$ defined by $f(x) = x^2$. The image of this set is the set of real numbers greater of or equal to zero. The reason is because x^2 is always greater or equal to zero for all real numbers x.

Definition 4.1.1

Let f: $A \longrightarrow B$ and g: $A \longrightarrow B$ be two mappings from set A to set B. These two mappings are said to be equal if they are equal on all element of A, i.e. f(a) = g(a) for all a ε A.

Definition 4.1.2

If f: $A \longrightarrow B$ and g:$B \longrightarrow C$ are mappings, let a ε A then f(a) ε B which is the domain of g. We can then obtain the image of f(a) under the mapping g to have gf(a) ε C. The map (gf)(a) = gf(a) is called the composition of f and g.

$$g\bullet f: A \longrightarrow C$$

where g•f(a) = g(f(a)).
We now show that composition of mapping satisfy the associative axiom.

Theorem 4.1.1

If $f: A \longrightarrow B$, $g: B \longrightarrow C$ and $h: C \longrightarrow D$ are mappings, then

$$(h\bullet g)\bullet f = h\bullet(g\bullet f).$$

Proof:

As remarked earlier, two mappings are equal if they are equal on every element of the domain. If a ε A then

$$((h\bullet g)\bullet f)(a) = (h\bullet g)(f(a) = h(g(f(a)))$$

and

$$h\bullet(g\bullet f)(a) = h(g\bullet f)(a) = h(g(f(a)))$$

Therefore

$$(h\bullet g)\bullet f)(a) = (h\bullet(g\bullet f))(a) \text{ for all } a \varepsilon A$$

then it follows that

$$(h\bullet g)\bullet f = h\bullet(g\bullet f)$$

Definition 4.1.3

Let $f: A \longrightarrow B$ be a mapping f is said to be one-to-one (or injective) if whenever x, y ε A and $x \neq y$ then $f(x) \neq f(y)$. In other words, f is one-to-one if f maps distinct elements of A to distinct elements of B.

Example 4.1.2

Consider the mapping $f: R \to R$ defined by $f(x) = x^2$.

This map is not injective because $f(-2) = 4 = f(2)$.

So f maps -2 and +2 to the same number 4.

Example 4.1.3

The mapping $f: R \longrightarrow R$ defined by $f(x) = x + 4$ is injective because if x, y ε R and $f(x) = f(y)$ then

$$x + 4 = y + 4$$

hence

$$x = y,$$

and so f is injective.

Definition 4.1.4

Let $f: A \longrightarrow B$ be a mapping, we say that f is an onto map (or surjective) if the image of f is the whole of B. In other words, f is surjective if given any b ε B there exists a ε A such that $f(a) = b$, i.e. $f(A) = B$.

Example 4.1.4

The mapping $f: R \longrightarrow$ defined by $f(x) = x^2$ is not onto.

For -9 ε R but there is no real number x ε R such that $f(x) = x^2 = -9$. In fact f(R) = all real numbers greater or equal to zero and so f(R) is not equal to the set of all real numbers hence f is not onto. However, the map $g: R \to R$ defined by $g(x) = x^3$ is onto.

Definition 4.1.5

Let A be a set, the identity map on A (denoted by I_A or simply id.) is the map defined by $I_A(a) = a$ for all a ε A. Identity map is both injective and surjective.

Definition 4.1.5

Let $f: A \longrightarrow B$ be a mapping, we say that f has an inverse if there exists a mapping $g: B \longrightarrow A$ such that $g \bullet f = I_A$ and $f \bullet g = I_B$. In this case we mean the composite maps $g \bullet f$ and $f \bullet g$ are the identity maps on A and B respectively.

Example 4.1.5

Let R^+ be set of real numbers which are greater or equal to zero, let f: $R^+ -> R^+$ be a map defined by $f(x) = x^2$ and consider another map g: $R^+ -> R^+$ defined by

$$g(x) = \sqrt{x}$$

then

$$f \bullet g(x) = x \text{ and } g.f(y) = y$$

then g is the inverse map of f and also f is the inverse map of g.

Theorem 4.1.2

Let $f : A \rightarrow B$ be a mapping which is injective and surjective then f has an inverse mapping.

Proof:

We need to define a map g:B $\longrightarrow$ A which is the inverse of f.
Since f is surjective, given any b ε B there exists a ε A such that f(a) = b and since f is injective, no other element of A is mapped onto b ε B. The element a ε A is therefore uniquely determined by b. We can then define g(b) = a hence by definition of g, we have

$$fg(b) = f(a) = b \text{ and } g\bullet f(a) = g(b) = a$$

for all b ε B. This shows that g is an inverse of f.

4.2 Linear Mapping

Definition 4.2.1

Let U and V be vector spaces over the same field K. A mapping f:U $\longrightarrow$ V is called a linear mapping if it satisfies the following two properties

(i) f(u + w) = f(u) + f(w) and
(ii) f(cu) = c f(u) for all u, w ε U and c ε K.

Recall that the two basic operations in a vector space are addition and scalar multiplication. The first property of a linear map shows that for a map to be a linear map, it must preserve addition while the second property shows that it must preserve scalar multiplication. The above two properties can be combined into one property, if a, b ε K and u, w ε U then f is linear if f(au + bw) = af(u) + bf(w).

A linear mapping is sometimes called linear transformation.

Example 4.2.1

Let f: $R^2 \rightarrow R^2$ be a mapping defined by f(x, y) = (2x + y, y) show that f is a linear map.

Solution

Suppose u, v ε R^2 where u = (a, b) and v = (x, y) then

$$u + v = (a, b) + (x, y) = (a + x, b + y)$$

$$f(u + v) = f(a + x, b + y) = (2(a + x) + b + y, b + y) \quad (2a + b + 2x + y, b + y)$$

but

$$f(u) + f(v) = f(a, b) + f(x, y) = (2a + b, b) + (2x + y, y) \quad (2a + b + 2x + y, b + y)$$

Hence

$$f(u + v) = f(u) + f(v)$$

If c ε K and v ε R^2 then

$$f(cv) = f(c(x, y) = f(cx, cy) \quad (2cx + cy, cy) = c(2x + y, y) = cf(v).$$

This shows that f is a linear map.

Example 4.2.2

Let V be a vector space of all n × n matrices over a field K. If A is a given matrix in V. Let f: V $\longrightarrow$ V be a mapping defined by f(M) = MA + AM. Show that f is a linear map.

Solution

If $M, N \varepsilon V$ and $c \varepsilon K$ then

$$f(M+N) = (M+N)A + A(M+N) = MA + NA + AM + AN$$
$$= (MA + AM) + (NA + AN) = f(M) + f(N)$$

and

$$f(cM) = (cM)A + A(cM) = c(MA) + c(AM) = c(MA + AM) = cf(M)$$

hence f is linear.

Definition: 4.2.2
A linear mapping $f: U \longrightarrow V$ which is also one-to-one is called a monomorphism while a linear mapping which is onto is called an epimorphism. A linear mapping which is both a monomorphism and an epimorphism is called an isomorphism.

Definition 4.2.3
Let $f: U \longrightarrow V$ be a linear map, the kernel of f denoted by Ker f is the set of all vectors in U that are mapped to zero vector in V, i.e. Ker (f) = $\{u \varepsilon U \mid f(u) = 0\}$ and the image of f denoted by Im(f) is the set of all vectors of the form f(u) ε V where u ε U, that is,

$$Im(f) = \{f(u) \mid u \varepsilon U\}$$

and this is also denoted by f(U).

Notice that if 0 is the zero vector then f(0) = 0. This is because if we take the scalar O (zero) then by axiom (ii) of the definition of linear map, we have

$$f(0) = f(0.v) = 0.f(v) = 0.$$

This shows that every linear mapping maps zero vector into the zero vector and so 0 ε Ker (f).

Theorem 4.2.1
Let $f: U \longrightarrow V$ be a linear mapping, the Kernel of f is a subspace of U and the image of f is a subspace of V.

Proof:
To show that Ker f is a subspace of U we show first that Ker f is non-empty then we show that if u, w ε Ker f and a, b ε K, then au + bw ε ker f. From the above remark, every linear maps takes zero vector to zero vector i.e. f(0) = 0 hence 0 ε Ker f and so Ker f ≠ φ. If u, w ε ker f then f(u) = 0 and f(w) = 0 then

$$f(au + bw) = af(u) + bf(w) = a.0 + b.0 = 0 + 0$$

hence au + bw ε ker f. This show that Ker f is a subspace of U.

Similarly since f(0) = 0 then 0 ε Im(f) and so Im(f) ≠ φ.
If $w_1, w_2 \varepsilon$ Imf and a, b ε K then there exists $u_1, u_2 \varepsilon U$ such that $f(u_1) = w_1$ and $f(u_2) = w_2$.

We then consider

$$f(au_1 + bu_2) = af(u_1) + bf(u_2) = aw_1 + bw_2$$

hence

$$a w_1 + b w_2 \varepsilon \operatorname{Im}(f).$$

Thus Im(f) is a subspace of W.

We have shown that given a linear map f: U $\longrightarrow$ V, the kernel of f is a subspace of U and the image of f is a subspace of V. The dimension of kernel of f is called the nullity of f while the dimension of the image of f is called the rank of f. We will prove a theorem which involves the nullity and rank of f, but before doing that we discuss the following.

Theorem 4.2.2

Let V be a vector space of dimension n and $\{u_1, u_2, \ldots, u_r\}$ a set of linearly independent vectors, this set can be extended to a basis of V.

Proof:

Let $\{v_1, v_2, \ldots, v_n\}$ be basis for V, consider the set $S = \{u_1, u_2, \ldots, u_r, v_1, v_2, \ldots, v_n\}$.

Since dimension of V is n then the set S cannot be linearly independent so the vectors are dependent and by theorem 1.3.1, one vector will be a linear combination of the preceding vectors. Since the u_is are linearly independent then no u_i is a linear combination of its predecessors. We then delete that vector which is a combination of its predecessor. Therefore no u_i will be deleted.

We continue to delete until we obtain the basis, which will now include every u_i.

Theorem 4.2.3

Let T : V $\longrightarrow$ U be a linear map then

$$\dim(U) = \text{Nullity}(T) + \text{Rank}(T).$$

Proof:

Recall that

$$\dim(\operatorname{Ker} T) = \text{Nullity}(T)$$

and

$$\dim(\text{Image } T) = \text{Rank}(T).$$

If dim (ker T) = m and dim(V) = n, let $\{v_1, v_2, \ldots, v_r\}$ be the basis for Ker(T) then by Theorem 5.2.2 this can be extended to basis $\{v_1, v_2, \ldots, v_{r-1}, v_r, v_{r+1}, \ldots, v_n\}$ of V. We show that $T(v_{r+1}), \ldots, T(v_n)$ is a basis for Im(T).

To show this, we need to show that they generate Im (T) and they are linearly independent.

Let u ε Im(T) then there exists v ε V such that T(v) = u but $\{v_1, \ldots, v_r, v_{r+1}, \ldots, V_n\}$ is a basis for V then

$$v = a_1 v_1 + a_2 v_2 + \ldots + a_n v_n.$$

and so

$$T(a_1 v_1 + a_2 v_2 + \ldots + a_n v_n) = u$$

$$\Rightarrow a_1 T(v_1) + \ldots + a_{r+1} T(v_{r+1}) + \ldots + a_n T(v_n) = u$$

but $T(v_1) = o, T(v_2) = o \ldots, T(v_r) = O$ since $v_1, v_2, \ldots, v_r$. is a basis for Ker(T) hence

$$a_{r+1} T(v_{r+1}) + \ldots + a_n T(v_n) = U,$$

shows that $T(v_{r+1}), T(v_{r+2}), \ldots, T(v_n)$ generate Im(T).

To show they are linearly independent, suppose $a_{r+1} T(v_{r+1}) + \ldots + a_n T(v_n) = O$ then

$$T(a_{r+1} v_{r+1} + \ldots + a_n v_n) = O$$

This implies $a_{r+1} v_{r+1} + \ldots + a_n v_n \,\varepsilon$ Ker(T).

But $\{v_1, v_2, \ldots, V_r\}$ generate Ker(T), this means that $a_{r+1}v_{r+1} + \ldots + a_n v_n$ is a linear combination of $v_1, v_2, \ldots, v_r$ and this is not possible because $(v_1, v_2, \ldots, v_r, v_{r+1}, \ldots v_n)$ are linearly independent hence

$$a_{r+1} = 0, a_{r+2} = 0, \ldots, a_n = 0.$$

It then follows that $T(v_{r+1}), T(v_{r+2}) \ldots, T(v_n)$ are linearly independent and so they form a basis for Im(T). Since $\{v_1, v_2, \ldots, v_r\}$ form a basis for Ker T then

$$\dim V = \text{Nullity}(T) + \text{Rank}(T).$$

Example 4.2.3

Consider the map $f: R^4 \to R^2$ defined by $f(x, y, z, w) = (2x + y + z + w, x + z - w)$. Find the basis and the dimension of (i) image of f and (ii) kernel of f.

Solution

The images of the usual basis on R^4 generate the image of f. The usual basis are $e_1 = (1, 0, 0, 0,), e_2 = (0, 1, 0, 0), e_3 = (0, 0, 1, 0)$,

$e_1 = (1, 0, 0, 0,), e_2 = (0, 1, 0, 0), e_3 = (0, 0, 1, 0)$ and $e_4 = (0, 0, 0, 1)$.

$$f(e_1) = f(1,0,0,0) = (2,1)$$

$$f(e_2) = f(0,1,0,0) = (1,0)$$

$$f(e_3) = f(0,0,1,0) = (1,1)$$

$$f(e_4) = f(0,0,0,1) = (1,-1)$$

We then form a matrix whose rows are the generators of image of f and reduce it to echelon form.

$$\begin{pmatrix} 2 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{pmatrix} \text{to} \begin{pmatrix} 2 & 1 \\ 0 & -1 \\ 0 & 1 \\ 0 & 2 \end{pmatrix} \text{to} \begin{pmatrix} 2 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Thus (2, 1), (0, -1) is a basis for image of f hence

$$\dim(\mathrm{Im}(f)) = 2 = \mathrm{rank}(f).$$

To determine the basis of kernel of f, we seek the set of all vectors $(x, y, z, w) \varepsilon R^4$ such that $f(x, y, z, w) = 0$ i.e.

$$(2x + y + z + w, x + z - w) = (0, 0)$$
$$\therefore 2x + y + z + w = 0 \text{ and } x + z - w = 0$$

Solving simultaneously, we have

$$x = w - z \text{ and } y = z - 3w$$

while z and w can be chosen arbitrarily, then the vector.

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} -z+w \\ z-3w \\ z \\ w \end{pmatrix} = Z\begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \end{pmatrix} + w\begin{pmatrix} 1 \\ -3 \\ 0 \\ 1 \end{pmatrix}.$$

The Kernel of f = nullity of f is the subspace of R^4 spanned by (-1, 1, 1, 0) and (1, -3, 0, 1) and so Nullity of f = 2.

4.3 Matrices and Linear Operators

Let V be a vector space over a field K. If dim(V) = n, let $\{e_1, e_2, \dots, e_n\}$ be a basis for V. If $v \varepsilon V$ then,

$$v = a_1 e_1 + a_2 e_2 + \dots + a_n e_n$$

where $a_i \varepsilon K$ are uniquely determined. We will call $(a_1, a_2, \dots a_n)$ the coordinate vector of v relative to $\{e_i\}$ and denote this by

$$[v]_e = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix}$$

where $e = (e_1, e_2, \dots, a_n)$.

Example 4.3.1

Recall that R^3 is a vector space over R. Vectors $\{e_1 = (1,1,1),\ e_2 = (1,1,0),\ e_3 = (1,0,0)\}$ form a basis for R^3. If $v = (10, 8, 3)$, then find $[v]_e$.

Solution

We need to write the vector v as a linear combination of e_1, e_2 and e_3 i.e. we need to find a, b, c such that

$$v = ae_1 + be_2 + ce_3 \text{ i.e.}$$

$$(10, 8, 3) = a(1, 1, 1) + b(1, 1, 0) + c(1, 0, 0) = (a + b + c, a + b, a)$$

hence

$$a + b + c = 10$$
$$a + b = 8$$
$$a = 3.$$

Solving simultaneously we obtain

$$a = 3, b = 5, c = 2,$$

hence

$$[v]_e = \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}.$$

Definition 4.3.1
Let V be a vector space over a field K, a linear operator on V is a linear map $T : V \longrightarrow V$ from V to itself. In this section, we will discuss how to represent a linear operator by a matrix given a particular basis. We will later extend this to any linear map.

Definition 4.3.2
Let $T : V \longrightarrow V$ be a linear operator on a vector space V and let $\{e_1, e_2, \ldots, e_n\}$ be a basis for V. Notice that the vectors $T(e_1), T(e_2), \ldots, T(e_n)$ are also in V and so each of them can be written as linear combinations of $(e_1, e_2, \ldots, e_n)$. hence

$$T(e_1) = a_{11}e_1 + a_{12}e_2 + a_{13}e_3 + \ldots + a_{1n}e_n$$
$$T(e_2) = a_{21}e_1 + a_{22}e_2 + a_{23}e_3 + \ldots + a_{2n}e_n$$
$$\vdots$$
$$T(e_n) = a_{n1}e_1 + a_{12}e_2 + a_{n3}e_3 + \ldots + a_{nn}e_n$$

we can simply write this in matrix form thus

$$\begin{pmatrix} T(e_1) \\ T(e_2) \\ \vdots \\ T(e_n) \end{pmatrix} = \begin{pmatrix} a_{11}\, a_{12}\, a_{13}, \cdots a_{1n} \\ a_{21}\, a_{22}\, a_{23} \cdots a_{2n} \\ \vdots \\ a_{n1}\, a_{n2}\, a_{n3} \cdots a_{nn} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

i.e.

$$T(e) = A\,e$$

where A is the matrix of coefficients. When we take the transpose of the matrix A of coefficients we get

$$\begin{pmatrix} a_{11}\ a_{21}\cdots a_{n1} \\ a_{12}\ a_{22}\cdots a_{n2} \\ \vdots \\ a_{1n}\ a_{2n}\cdots a_{nn} \end{pmatrix}$$

denoted by $[T]_e$ and $[T]_e$ is called the matrix representing T in the basis $e = (e_1, e_2, \ldots, e_n)$ or the matrix of T relative to the basis $(e_1, e_2, \ldots, e_n)$.

Example 4.3.2

Let $T : R^2 -> R^2$ be a linear mapping defined by T(x, y) = (5x + y, 3x - 2y). Find the matrix $\{T\}_u$ in the basis $u = (1, 2)$ and $(2, 3)$ of R^2.

Solution

$$T(1, 2) = (5.1 + 2, 3.1 - 2.2) = (7, -1)$$
$$(7, -1) = a(1, 2) + b(2, 3) = (a + 2b, 2a + 3b)$$
$$a + 2b = 7$$
$$2a + 3b = -1,$$

solving simultaneously we have

$$a = -23 \text{ and } b = 15$$

Also

$$T(2, 3) = (5.2 + 3, 3.2 - 2.3) = (13, 0)$$
$$(13, 0) = c(1, 2) + d(2, 3) = (c + 2d, 2c + 3d)$$

so

$$c + 2d = 13$$
$$2c + 3d = 0.$$

Solving simultaneously, we have

$$c = -39 \text{ and } d = 26.$$

The matrix of coefficient is

$$\begin{pmatrix} -23 & 15 \\ -39 & 26 \end{pmatrix}$$

and so

$$[T]_u = \begin{pmatrix} -23 & -39 \\ 15 & 26 \end{pmatrix}.$$

Example 4.3.3

Let $T : R^3 \rightarrow R^3$ be a linear operator defined by T(x, y, z) = (2y + z, x - 4y, 3x),

(i) find the matrix $[T]_e$ of T in the basis

$$e = \{e_1 = (1, 1, 0), e_2 = (1, 0, 0), e_3 = (0, 0, 1)\}.$$

(ii) If v = (3, 5, -2), verify that $[T]_e[v]_e = [T(v)]_e$.

Solution

(i) $T(e_1) = T(1,1,0) = (2,-3,3)$

$$= a(1,1,0) + b(1,0,0) + c(0,0,1) = (a+b, a, c)$$

hence

$$a+b=2$$
$$a=-3$$
$$c=3$$
$$\therefore b=5$$

Therefore

$$T(e_1) = -3e_1 + 5e_2 + 3e_3.$$

$T(e_2) = T(1,0,0) = (0,1,3) = a(1,1,0) + b(1,0,0) + c(0,0,1) = (a+b, a, c)$

hence

$$a+b=0$$
$$a=1$$
$$c=3$$
$$\therefore b=-1$$

Therefore

$$T(e_2) = (1,1,0) - (1,0,0) + 3(0,0,1)$$

$T(e_3) = T(0,0,1) = (1,0,0) = a(1,1,0) + b(1,0,0) + c(0,0,1) = (a+b, a, c)$

$$a+b=1$$
$$a=0$$
$$c=0.$$

$$\therefore T(e_3) = 0(1,1,0) + (1,0,0) + 0(0,0,1)$$

$$\therefore [T]_e = \begin{pmatrix} -3 & 5 & 3 \\ 1 & -1 & 3 \\ 0 & 1 & 0 \end{pmatrix}^t = \begin{pmatrix} -3 & 1 & 0 \\ 5 & -1 & 1 \\ 3 & 3 & 0 \end{pmatrix}$$

(ii) Since $v = (3, 5, -2)$ then

$$v = a\,e_1 + b\,e_2 + c\,e_3.$$

$$(3,5,-2) = a(1,1,0) + b(1,0,0) + c(0,0,1) = (a+b, a, c)$$

$$\therefore a+b=3$$
$$a=5$$

hence

$$b=-2$$
$$c=-2$$

So

$$[v]_e = \begin{pmatrix} 5 \\ -2 \\ -2 \end{pmatrix}$$

$$T(v) = T(3,5,-2) = (2.5 - 2, 3 - 4.5, 3.3) = (8,-17,9)$$

hence

$$(8,-17,9) = a(1,1,0) + b(1,0,0) + c(0,0,1) = (a+b, a, c)$$

therefore

$$a+b=8,\ a=-17 \text{ and } c=9,$$

$$b=25.$$

It follows that

$$[T(v)]_e = \begin{pmatrix} -17 \\ 25 \\ 9 \end{pmatrix}$$

but

$$[T]_e[v]_e = \begin{pmatrix} -3 & 1 & 0 \\ 5 & -1 & 1 \\ 3 & 3 & 0 \end{pmatrix}\begin{pmatrix} 5 \\ -2 \\ -2 \end{pmatrix} = \begin{pmatrix} -17 \\ 25 \\ 9 \end{pmatrix} = [T(v)]_e$$

Therefore

$$[T]_e[v]_e = [T(v)]_e.$$

We have shown that given a linear operator $T : V \longrightarrow V$ and a basis $e = (e_1, e_2, ..., e_n)$ we can construct a matrix $[T]_e$ which represents the linear operator in the basis e. The question that can readily come to mind is that if we have another basis $u = (u_1, u_2, u_3, ..., u_n)$ then we can get another matrix $[T]_u$ of the same linear operator T in the basis u, then what will be the relationship between $[T]_e$ and $[T]_u$. This leads us to the following definition.

Definition 4.3.3

Let V be a vector space over a field K, let $\{e_1, e_2, ..., e_n\}$ and $\{u_1, u_2, ..., u_n\}$ be basis for V. Since $\{e_1 ... e_n\}$ is a basis and $(u_1, ..., u_n)$ are in particular, vectors in V then each of the u_is can be written uniquely as a linear combination of the $\{e_i\}$.

$$u_1 = a_{11}e_1 + a_{12}e_2 + ... + a_{1n}e_n$$

$$u_2 = a_{21}e_1 + a_{22}e_2 + ... + a_{2n}e_n$$

$$\vdots$$

$$u_n = a_{n1}e_1 + a_{n2}e_2 + ... + a_{nn}e_n$$

The transpose P of the above matrix of coefficients is called the transition matrix from basis $\{e_i\}$ to basis $\{u_i\}$.

$$P = \begin{pmatrix} a_{11} & a_{21} & \cdot & \cdot & \cdot & a_{n1} \\ a_{12} & a_{22} & \cdot & \cdot & \cdot & a_{n2} \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ a_{1n} & a_{2n} & \cdot & \cdot & \cdot & a_{nn} \end{pmatrix}$$

Since the vectors $u_1, u_2, ..., u_n$ are linearly independent then the matrix P is invertible and P^{-1} is the transition matrix from basis $\{u_i\}$ to basis e_i.

Example 4.3.4

Consider the two basis of the vector space R^3,
$e = \{ e_1 = (1,1,1), e_2 = (1,1,0), e_3 = (1,0,0) \}$ and
$u = \{ u_1 = (2,3,4), u_2 = (3,0,2), u_3 = (3,1,4) \}$. Find the transition matrix from $\{ e_i \}$ to $\{ u_i \}$.

Solution

$$u_1 = (2, 3, 4) = a(1, 1, 1) + b(1, 1, 0) + c(1, 0, 0) = (a + b + c, a + b, a)$$

then

$$a + b + c = 2$$
$$a + b = 3$$
$$a = 4$$

then

$$a = 4, b = -1 \text{ and } c = -1.$$

Also

$$u_2 = (3, 0, 2) = a(1,1,1) + b(1,1,0) + c(1,0,0) = (a + b + c, a + b, a)$$
$$a = 2, a + b = 0, a + b + c = 3$$

hence

$$a = 2, b = -2, c = 3$$
$$= (3, 1, 4) = a(1,1,1) + b(1,1,0) + c(1,0,0) = (a + b + c, a + b, a)$$
$$a = 4, a + b = 1, a + b + c = 3$$

hence

$$a = 4, b = -3, c = 2$$

matrix of coefficient is $\begin{pmatrix} 4 & -1 & -1 \\ 2 & -2 & 3 \\ 4 & -3 & 2 \end{pmatrix}$

hence transition matrix from $\{e_i\}$ to $\{u_i\}$ is

$$P = \begin{pmatrix} 4 & 2 & 4 \\ -1 & -2 & -3 \\ -1 & 3 & 2 \end{pmatrix}$$

A transition matrix P from the basis $e = \{ e_i \}$ to the basis $\{u_i\}$, its effect is to transform the coordinates of a vector in the basis $\{u_i\}$ back to the coordinates in the basis $\{ e_i \}$. This means that if $v \in V$ then

$$P[v]_u = [v]_e$$

For instance in Example 4.3.4 consider the vector $v = (3, 5, -2)$, we want to calculate $[v]_e$ and $[v]_u$.

$$(3, 5, -2) = a(1, 1, 1) + b(1, 1, 0) + c(1, 0, 0) = (a + b + c, a + b, a)$$
$$\therefore a = -2, b = 7, c = -2.$$

Therefore,

$$[v]_e = \begin{pmatrix} -2 \\ 7 \\ -2 \end{pmatrix}$$

Also to calculate $[v]_u$,

$$(3, 5, -2) = a(2, 3, 4) + b(3, 0, 2) + c(3, 1, 4) = (2a + 3b + 3c, 3a + c, 4a + 2b + 4c)$$

$$2a + 3b + 3c = 3$$

$$3a + c = 5$$

$$4a + 2b + 4c = -2$$

Solving the above equations simultaneously we obtain

$$a = \frac{42}{10}, b = \frac{58}{10} \text{ and } c = \frac{-76}{10}$$

hence

$$[v]_u = \begin{pmatrix} \frac{42}{10} \\ \frac{58}{10} \\ \frac{-76}{10} \end{pmatrix}$$

We therefore have

$$P[V]_u = \begin{pmatrix} 4 & 2 & 4 \\ -1 & -2 & -3 \\ -1 & 3 & 2 \end{pmatrix} \begin{pmatrix} \frac{42}{10} \\ \frac{58}{10} \\ \frac{-76}{10} \end{pmatrix} = \begin{pmatrix} -2 \\ 7 \\ 2 \end{pmatrix} = [V]_e$$

Let V be a vector space over a field K if $e = \{e_1, e_2, \ldots e_n\}$ and $u = \{u_1, u_2, \ldots, u_n\}$ are two basis of the vector space.

If T: V $\longrightarrow$ V is a linear operator, as we discussed earlier, we can construct matrices $[T]_e$ and $[T]_u$ which represent the linear operator T. The next theorem establish the relationship between such matrices.

Theorem 4.3.1

If $T : V \longrightarrow V$ is a linear operator and V a vector space with basis $e = (e_1, e_2, \ldots, e_n)$ and $u = (u_1, u_2, \ldots, u_n)$. If P is a transition matrix from basis $\{e_i\}$ to basis $\{u_i\}$ then

Proof:

Let $v \in V$ be any vector then since P is the transition matrix from $\{e_i\}$ to $\{u_i\}$ then $P[v]_u = [v]_e$ then

$$P^{-1}[T]_e P[u]_u = P^{-1}[T]_e[v]_e = P^{-1}[T(v)]_e$$

since $P[v]_u = [v]_e$ then

$$[v]_u = P^{-1}[v]_e .$$

It then follows that

$$P^{-1}[T(v)]_e = [T(v)]_u = [T]_u[v]_u ,$$

hence

$$P^{-1}[T]_e P[v]_u = [T]_u[v]_u ,$$

since this is true for any vector v ε V, it follows that

$$P^{-1}[T]_e P = [T]_u .$$

Definition 4.3.4
Let A and B be n × n matrices if there exists an invertible matrix P such that $B = P^{-1}AP$ then we say that B is similar to A. From definition 4.3.4 and theorem 4.3.1,it then follows that two matrices representing the same linear operator are similar to each other.
Example 4.3.5

Consider a linear operator T: V → V defined by T(x, y) = (2x - 3y, x + y).
If $e = \{e_1 = (1,2), e_2 = (2,3)\}$ and $u = \{u_1 = (1,3), u_2 = (1,4)\}$. Find the transition matrix P from $\{e_i\}$ to $\{u_i\}$ and show that $P^{-1}[T]_e P = \{T\}_u$.

Solution

We find $P, [T]_e$ and $[T]_u$, to find P, we write each vector u_i as a linear combination of e_i. $u_i = (1,3) = a(1,2) + b(2,3) = (a + 2b, 2a + 3b)$

hence

$$a + 2b = 1$$
$$2a + 3b = 3$$

hence

$$a = 3,\ b = -1$$

$$u_2 = (1,4) = a(1,2) + b(2,3) = (a + 2b, 2a + 3b)$$
$$a + 2b = 1$$
$$2a + 3b = 4$$

Solving simultaneously, we have

$$a = 5 \text{ and } b = -2$$

Matrix of coefficient is

$$\begin{pmatrix} 3 & -1 \\ 5 & -2 \end{pmatrix}$$

hence

$$P = \begin{pmatrix} 3 & 5 \\ -1 & -2 \end{pmatrix}$$

We now calculate $[T]_e$

$$T(e_1) = T(1,2) = (-4,3) = a(1,2) + b(2,3) = (a + 2b, 2a + 3b)$$

hence

$$a + 2b = -4$$
$$2a + 3b = 3$$

hence $a = 18$, $b = -11$

$$T(e_2) = T(2, 3) = (-5, 5) = a(1, 2) + b(2, 3) = (a + 2b, 2a + 3b)$$

$$a + 2b = -5$$
$$2a + 3b = 5,$$

hence

$$a = 25, \ b = -15$$

hence

$$T(e_1) = 18(1, 2) - 11(2, 3)$$
$$T(e_2) = 25(1, 2) - 15(2, 3)$$

$$[T]_e = \begin{pmatrix} 18 & 25 \\ -11 & -15 \end{pmatrix}$$

Similarly, we calculate $[T]_u$

$$T(u_1) = T(1, 3) = (-7, 4) = a(1, 3) + b(1, 4) = (a + b, 3a + 4b)$$

hence

$$a + b = -7$$
$$3a + 4b = 4$$

hence
$a = -32$ and $b = 25$

$$T(u_2) = T(1, 4) = (-10, 5) = a(1, 3) + b(1, 4) = (a + b, 3a + 4b)$$

$$a + b = -10$$

hence

$$a = -45$$
$$3a + 4b = 5 \ b = 35$$

We therefore have

$$T(u_1) = -32(1, 3) + 25(1, 4)$$
$$T(u_2) = -45(1, 3) + 35(1, 4)$$

hence

$$[T]_u = \begin{pmatrix} -32 & -45 \\ 25 & 35 \end{pmatrix}$$

$$P^{-1} = \begin{pmatrix} 2 & 5 \\ -1 & -3 \end{pmatrix}$$

$$P^{-1}\,[T]_e\,P = \begin{pmatrix} 2 & 5 \\ -1 & -3 \end{pmatrix} \begin{pmatrix} 18 & 25 \\ -11 & -15 \end{pmatrix} \begin{pmatrix} 3 & 5 \\ -1 & -2 \end{pmatrix}$$

$$= \begin{pmatrix} -19 & -25 \\ 15 & 20 \end{pmatrix} \begin{pmatrix} 3 & 5 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} -32 & -45 \\ 25 & 35 \end{pmatrix} = [T]_u$$

Example 4.3.6
Let V a vector space of $(n \times n)$ matrices over a field K. Show that similarity of $n \times n$ matrices is an equivalence relation on V.

Solution

We need to show that similarity of matrices is reflexive, symmetric and transitive. If $A \varepsilon V$, notice that the identity matrix I is invertible hence

$$A = I^{-1} A I.$$

This shows that every matrix is similar to itself hence reflexive property is proved. Suppose A and B are $n \times n$ matrices such that A is similar to B, we need to show that B is also similar A. Since A is similar to B then there exists an invertible matrix P such that $A = P^{-1} B P$, we then multiply both sides of the equation on the left by P and on the right by P^{-1} to have

$$P A P^{-1} = PP^{-1} B PP^{-1}, (\text{since } PP^{-1} = I) = IBI = B.$$

If we let $P^{-1} = Q$ then $P = (P^{-1})^{-1} = Q^{-1}$, hence

$$Q^{-1} A Q = P A P^{-1} = B.$$

Since P is invertible, $P^{-1} = Q$ is also invertible and so we have found an invertible matrix Q such that $B = Q^{-1} A Q$ hence B is similar to A and symmetric property is proved. Suppose A, B and C are $n \times n$ - matrices such that A is similar to B and B is similar to C then there exists invertible matrices P and R such that

$$A = P^{-1} B P \text{ and } B = R^{-1} C R.$$

We need to show that A is similar to C. We substitute for B in $P^{-1} B P$ to have

$$P^{-1} B P = P^{-1} R^{-1} C R P = (R P)^{-1} C R P.$$

If $Q = RP$ then we have

$$A = P^{-1} BP = (RP)^{-1} CRP = Q^{-1} C Q.$$

We have found a matrix $Q = RP$ such that $A = Q^{-1} C Q$ hence A is similar to C and so transitivity is established. This thus shows that similarity of matrices is an equivalence relation.

We have been discussing how we can represent a linear operator on a vector space V by a matrix when given a basis. We now consider the general case of a linear map from a vector space V to a vector space U.

Let $f: V \longrightarrow U$ be a linear map from vector space V where $\dim V = m$ to a vector space U where $\dim U = m$. Let $\{e_1, e_2, \ldots, e_n\}$, be a basis for V and $\{u_1, u_2, \ldots u_m\}$ a basis for U. The vectors $f(e_1), f(e_3), \ldots, f(e_n)$ are elements in U and since $\{u_1, u_2, \ldots, u_m\}$ is a basis for U then each of the vectors $f(e_1), f(e_2) \ldots f(e_n)$ can be written as linear combination of the

$\{u_i\}$.

$$f(e_2) = a_{21}u_1 + a_{22}u_2 + \ldots + a_{2m}u_m$$

$$\vdots$$

$$f(e_n) = a_{n1}u_1 + a_{n2}u_2 + \ldots + a_{nm}u_m$$

The matrix of coefficients of the above equations iF

$$\begin{pmatrix} a_{11}\,a_{12}\cdots a_{1m} \\ a_{21}\,a_{22}\cdots a_{2m} \\ \vdots \\ \vdots \\ \vdots \\ a_{n1}\,a_{n2}\cdots a_{nm} \end{pmatrix}$$

The transpose of this matrix denoted by

$$[f]^u_e = \begin{pmatrix} a_{11}\,a_{21}\cdots a_{n1} \\ a_{12}\,a_{22}\cdots a_{n2} \\ \cdots \\ \cdots \\ \cdots \\ a_{1m}\,a_{2m}\,a_{nm} \end{pmatrix}$$

is the matrix representation of f relative to the basis $\{e_i\}$ $\{u_i\}$ and it is an $m \times n$ matrix.

Example 4.3.7

Let $f: R^2 -> R^3$ be a linear map defined by $f(x, y) = (x - y, 2x, x + 2y)$

(i) Find the matrix of f in the following basis of R^2 and R^3

$\{u_1 = (1, 3), u_2 = (2, 5)\}$ and $\{v_1 = (1,1,1), v_2 = (1,1,0)\ v_3 = \{1,0,0)$

(ii) If $w = (x, y)$ is any vector in R^2 verify that $[f]^v_e[w]_u = [f(w)]_v$

Solution

$$f(u) = f(1, 3) = (-2, 2, 7) = a(1, 1, 1) + b(1, 1, 0) + c(1, 0, 0)$$
$$= (a + b + c, a + b, a)$$
$$a = 7$$
$$a + b = 2 \text{ then } b = -5$$
$$a + b + c = -2 \text{ and } c = -4$$

$$f(u_2) = f(2, 5) = (-3, 4, 12) = a(1, 1, 1) + b(1, 1, 0) + c(1, 0, 0)$$
$$= (a + b + c, a + b, a)$$
$$a = 12$$
$$a + b = 4 \text{ hence } b = -8$$
$$a + b + c = -3 \text{ and } c = -7$$

This shows

$$f(u_1) = 7(1,1,1) - 5(1,1,0) + -4(1,0,0)$$
$$f(u_2) = 12(1,1,1) - 8(1,1,0) - 7(1,0,0)$$

hence

$$[f]_u^v = \begin{pmatrix} 7 & 12 \\ -5 & -8 \\ -4 & -7 \end{pmatrix}$$

For a vector

$$w = (x, y) = a(1, 3) + b(2, 5) = (a + 2b, 3a + 5b$$
$$a + 2b = x$$
$$3a + 5b = y$$

Solving simultaneously, we have

$$a = -5x + 2y \text{ and } b = 3x - y.$$

This shows that

$$[w]_u = \begin{pmatrix} -5x + 2y \\ 3x - y \end{pmatrix}$$

$$f(w) = f(x, y) = (x - y, 2x, x + 2y)$$
$$= a(1, 1, 1) + b(1, 1, 0) + c(1, 0, 0)$$
$$= (a + b + c, a + b, a)$$
$$a = x + 2y$$
$$a + b = 2x$$
$$a + b + c = x - y$$

On solving simultaneously, we have $b = x - 2y$ and $c = -x, -y$, hence

$$[f(w)]_v = \begin{pmatrix} x + 2y \\ x - 2y \\ -x - y \end{pmatrix}$$

$$[f]_u^v[w]_u = \begin{pmatrix} 7 & 12 \\ -5 & -8 \\ -4 & -7 \end{pmatrix} \begin{pmatrix} -5x + 2y \\ 3x - y \end{pmatrix} = \begin{pmatrix} -35x + 14y + 36x - 12y \\ 25x - 10y + 24x + 8y \\ 20x - 8y - 21x + 2y \end{pmatrix}$$

$$= \begin{pmatrix} x + 2y \\ x - 2y \\ -x - y \end{pmatrix} = [f(w)]_v$$

4.4 Trace

Definition 4.4.1

Let $A = (a_{ij})$ be an $n \times n$ matrix the trace of A (denoted by tr(A)) is the sum of the diagonal entries of A, that is,

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

Example 4.4.1

Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$

$$\text{Trace of } A = \operatorname{tr}(A) = 1 + 5 + 9 = 15$$

Theorem 4.4.1
Let A and B be n × n matrices where $A = (a_{ij})$ and $B = (b_{ij})$ then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Proof:
Recall that since $A = (a_{ij})$ and $B = (b_{ij})$ then

$$AB = (C_{ij}) \text{ where } c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$$

hence

$$\operatorname{tr}(AB) = \sum_{i=1}^{n} C_{ii} = \sum_{i=1} \sum_{j=1}^{n} (a_{ij} b_{ji}).$$

Also $BA = d_{kj}$ where $d_{kj} = \sum_{i=1}^{n} b_{ki} a_{ij}$

$$\operatorname{tr}(BA) = \sum_{j=1}^{n} d_{jj} = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} b_{ji} a_{ij} \right)$$

but $\sum_{j=1}^{n} \left(\sum_{i=1}^{n} b_{ji} a_{ij} \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} b_{ji} \right) = \operatorname{tr}(AB)$

From theorem 4.4.1 we can deduce the following.

Theorem 4.4.2
If A and B are n × n matrices, if A is similar to B then $\operatorname{tr}(A) = \operatorname{tr}(B)$.

Proof:
Since A is similar to B then there exists invertible matrix P such that $A = P^{-1}BP$. From theorem 4.4.1 then

$$\operatorname{tr}(A) = \operatorname{tr}(P^{-1}BP) = \operatorname{tr}(PP^{-1}B)$$
$$\operatorname{tr}(IB) = \operatorname{tr}(B).$$

From theorem 4.4.2, we can then define the trace of a linear operator.

Definition 4.4.2
Let $T : V \longrightarrow V$ be a linear operator on a finite dimensional vector space V, the trace of T is the trace of a matrix representing T in any basis. This definition does not depend on the basis of the vector space V since if $e = \{e_i\}$ and $u = \{u_i\}$ are two basis of V then the matrices $[T]_e$ and $[T]_u$ representing T in the basis are similar and by theorem 4.4.2, they have the same trace.

Exercise 4.5

1. Let U, V and W be a vector spaces over a field K and $f: U \longrightarrow V$, $g: V \longrightarrow W$ is a linear mappings show that the composition $g.f: U \longrightarrow W$ is a linear mapping.
2. Let V be the vector space of all $n \times n$ matrices over a field K and $f: V \longrightarrow V$ a map defined by $f(A) = AM - MA$ where M is a fixed $n \times n$ matrix and A any element in V. Show that f is a linear map.
3. Let V and W be vector spaces over a field K and let $f: V \longrightarrow W$, $g: V \longrightarrow W$ be linear mappings from V to W. If we define the map $(f+g)(v) = f(v) + g(v)$ for all $v \varepsilon V$, show that f + g is also a linear map.
4. Let $T: R^3 \to R^3$ be a linear mapping defined by $T(x, y, z) = (x - y + 2z, -x, -2y + z, y + z)$, find a basis and the dimension of the (i) image of T and (ii) kernel of T.
5. Let $T: V \longrightarrow W$ be a linear map which is one-to one. If $v_1, v_2, \ldots, v_n$ are linearly independent vectors in V, show that $T(v_1), T(v_2), \ldots, T(v_n)$ are also linearly independent in W.
6. If U and V are vector spaces over a field K and $f: U \longrightarrow V$ is a linear map, if $c \varepsilon K$ such that $c \neq 0$, show that f and cf have the same image and the same kernel.
7. If $f: R^4 \to R^3$ is defined by $f(x,y,z,w) = (-2x-5z+3w, x+2z-w, 2x+3y-z+w)$. Find the dimension and basis for Im(f) and ker(f).
8. Let $T: R^3 \to R^3$ be a mapping defined by $T(x, y) = (5x - 2y, 3x + y)$
   (i) Show T is a linear operator.
   (ii) Compute the matrix representing T in the basis $e = \{1, 4), (2, 7)\}$ and in the basis $u = \{(1, 1), (0, 1)\}$. $a(1,4)+b(2,7) = 1,1$
   (iii) If $v = (x, y) \varepsilon R^2$, show that $[T]_e [v]_e = [T(v)]_e$.

Same as (9)

9. Let $f: R^3 \to R^2$ be a linear map defined by $f(x, y, z) = (2x - y + 4z, 3x + 2y + z)$. Find the matrix of f in the following basis of R^3 and R^2
   $\{u_1 = (1,1,1), u_2 = (1,1,0), u_3 = (1,0,0)\}$ $\{e_1 = (1,1), e_2 = (0,1)\}$.
10. If $T: R^3 \to R^3$ is a linear operator defined by $T(x, y, z) = (3x-2y+z, 5y-2z, 9x-z)$. Find tr(T).
11. Let V be a vector space of 2×2 matrices over set R of real numbers. Let $A = \begin{pmatrix} 3 & 4 \\ 2 & 1 \end{pmatrix}$ be a fixed element of V. We define a linear operator $T: V \longrightarrow V$ by $T(M) = AM + MA$. Find the matrix of T in the basis
   $$\{e_1 = (\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} e_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$
12. If A is an invertible matrix and it is similar to matrix B then show that B is also invertible.
13. Let $T: U \longrightarrow V$ be a linear map. If $u_1, u_2, \ldots, u_n$ <- V is such that $T(u_1), T(u_2), \ldots, T(u_n)$ are linearly independent, then show that $u_1, u_2, \ldots, u_n$ are also linearly independnent.
14. Show that the following maps are not linear.
   (a) $T: R^3 \longrightarrow R^2$ defined by $T(a, b c) = (a + 1, b + c)$
   (b) $T: R^2 \longrightarrow R^2$ defined by $T(a, b) = (a^2, b^2)$
15. Let $T: R^3 \longrightarrow R^3$ be defined by $T(a, b, c) = (a - 3b - 2z, b - 4c, c)$ show that T is an invertible operator. Find T^{-1}.

CHAPTER FIVE

110313091

DETERMINANTS

5.1 DEFINITION AND EXAMPLES

Let A be an n × n matrix, over a field K, the determinant of A (denoted by det(A) or $|A|$ is a specific scalar assigned to A. The determinant is therefore a function which assigns to every square matrix, a scalar obtained from that matrix. Determinant is an important tool in computing the inverse of an invertible matrix and in solving system of linear equations. We will first give the details of method of computing the determinant of 2 × 2 and 3 × 3 matrices. Suppose A is a 1 × 1 matrix i.e. A = (a)

Then

$$|A| = a.$$

If A is a 2 × 2 matrix i.e.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

then

$$|A| = a_{11}a_{22} - a_{21}a_{12}.$$

This simply implies that if A is a 2 × 2 matrix we obtain the determinant by taking the product of the two diagonal entries and subtracting the product of the two other entries.

Example 5.1.1

If $A = \begin{pmatrix} 3 & 2 \\ -1 & 3 \end{pmatrix}$, find $|A|$.

Solution

$$\begin{vmatrix} 3 & 2 \\ -1 & 3 \end{vmatrix} = 3.3 - (-1).2 = 9 + 2 = 11$$

Let A be a 3 × 3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

The method we use to evaluate $|A|$ is to reduce it to a linear combination of the determinants of 2 × 2 matrices.

To do this, we can use any row or any column. We will use the first row vector (a_{11}, a_{12}, a_{13}) to calculate the determinant. We begin with the first entry in that row which is a_{11} and delete the row and the column which the entry belongs i.e. we delete the first row and first column of matrix A.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

We are then left with a 2 × 2 matrix

$$\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}$$

$$\text{the determinant of this matrix} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$$

is called the minor of a_{11}, we similarly obtain the minor for a_{12} by deleting the first row and second column of A.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

hence

$$\text{minor for } a_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$$

we similarly obtain

$$\text{the minor for } a_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

We define the co-factor A_{ij} of element a_{ij} as plus or minus the minor of a_{ij} according as $(i+j)$ is even or odd. If M_{ij} represents minor of a_{ij} The co-factor of $a_{ij} = A_{ij} = (-1)^{i+j} M_{ij}$

In the 3 × 3 matrix above,

$$A_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = a_{22}\, a_{33} - a_{32}\, a_{23}$$

$$A_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = -(a_{21}\, a_{33} - a_{31}\, a_{23})$$

$$A_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} = (a_{21}\, a_{32} - a_{31}\, a_{22})$$

The determinant of the 3 × 3 matrix A is the linear combination of these three co-factors whose coefficients is the first row of A

$$|A| = a_{11}\,(a_{22} \cdot a_{33} - a_{32}\, a_{23}) - a_{12}\,(a_{21}\, a_{33} - a_{31}\, a_{23}) + a_{13}\,(a_{21}\, a_{32} - a_{31}\, a_{22})$$

Example 5.1.2

If $A = \begin{pmatrix} 1 & 4 & 0 \\ 2 & 5 & 2 \\ 3 & -1 & 3 \end{pmatrix}$, calculate $|A|$.

Solution

We use the first row to calculate the determinant.

$$A_{12} = \begin{vmatrix} 5 & 2 \\ -1 & 3 \end{vmatrix} = 5(3) - (-1)2 = 15 + 2 = 17$$

$$A_{12} = \begin{vmatrix} 2 & 2 \\ 3 & 3 \end{vmatrix} = -(2(3) - (3)2) = 6 - 6 = 0$$

$$A_{13} = \begin{vmatrix} 2 & 5 \\ 3 & -1 \end{vmatrix} = 2(-1) - 3(5) = -2 - 15 = -17$$

hence

$$|A| = 1(17) - 4(0) + 0(-17) = 17.$$

The signs to be assigned in front of the minors can best be remembered by the array of alternating signs.

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$$

for the 3 × 3 matrix. Generally the sign starts with + on first row first column and changes sign for each down-ward step or side-ways step.

We wish to show that to compute the determinant of a matrix, any row or any column can be used. Recall that if we use first row of matrix A then

$$|A| = a_{11}(a_{22}a_{33} - a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22})$$

$$= a_{11}a_{12}a_{33} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} + a_{12}a_{31}a_{23} + a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22}.$$

Suppose we now use the third column of matrix A to compute the determinant, then the co-factor A_{13} will be obtained from

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

to have

$$\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

hence

$$A_{13} = +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Similarly

$$A_{23} = -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$$

and

$$A_{33} = +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

$$\therefore |A| = a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} - a_{23}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} + a_{33}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

$$= a_{13}(a_{21}a_{32} - a_{22}a_{31}) - a_{23}(a_{11}a_{32} - a_{12}a_{31}) + a_{33}(a_{11}a_{22} - a_{12}a_{21})$$

$$= a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{23}a_{11}a_{32} + a_{23}a_{12}a_{31}a_{23}a_{12}a_{31} + a_{33}a_{11}a_{22} - a_{33}a_{21}a_{12}.$$

This gives exactly the same result as when we computed $|A|$ using the first row. This shows that in computing the determinant of any 3 × 3-matrix, any row or column can be used, the same result will be obtained.

If A is an n × n-matrix, if n is a large number (in fact if $n>3$) and we need to compute $|A|$, then if we use the method just described, the number of terms involved in the process becomes very large and cumbersome. We then need to develop a easier and more efficient method of evaluating such determinants. To do this we need to show some properties of determinants which will be applied in this method.

5.2 Properties of Determinants and Determinant of Matrices of Large Size

We will discuss some basic properties of determinants and illustrate these properties using 3 × 3 matrix.

1. If A is any matrix and we interchange two rows (or two columns) to get another matrix B, then

$$|A| = -|B|.$$

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

If we interchange the second and third row then

$$B = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

Using the first row to evaluate our matrix, we have

$$|A| = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$|B| = a_{11}\begin{vmatrix} a_{32} & a_{33} \\ a_{22} & a_{23} \end{vmatrix} - a_{12}\begin{vmatrix} a_{31} & a_{33} \\ a_{21} & a_{23} \end{vmatrix} + a_{13}\begin{vmatrix} a_{31} & a_{32} \\ a_{21} & a_{22} \end{vmatrix}$$

When we examine the above two determinants closely, we see that $|A| = -|B|$. This shows that the determinant of a matrix changes sign when two rows (or two columns) are interchanged. From this property, a very simple property of determinant can be deduced.

2. If A is a square matrix in which two rows (or two columns) are identical then $|A| = 0$. To show this, suppose you interchange those two identical rows, we will still obtain the same matrix A. By property (1) then

$$|A| = -|A|$$

which implies $2|A| = 0$, hence

$$|A| = 0.$$

3. If A has a row (or column) of zeros then $|A| = 0$.
This property is easy to see because since we can use any row or column to compute the determinant then we use the row in which all the entries is zero to compute the determinant then $|A| = 0$.

4. If A is a square matrix and we add a scalar multiple of one row to another row to get a matrix B, then $|A| = |B|$.

Let $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ *and let c be a scalar such that we add the multiple of second row with c to first row and replace the first with result i.e.* $R_1 \to cR_2 + R_1$ *to have*

$$B = \begin{pmatrix} a_{11}+ca_{21} & a_{12}+ca_{22} & a_{13}+ca_{32} \\ a_{21} & a_{22} & a_{23} \\ a_{23} & a_{32} & a_{33} \end{pmatrix}$$

Using first row to compute the determinant, we have

$$|B| = (a_{11}+ca_{21})\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - (a_{12}+ca_{22})\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + (a_{13}+ca_{23})\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + ca_{21}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$$

$$- ca_{22}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} + ca_{23}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$+ c\left(a_{21}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{22}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{23}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}\right)$$

In above expression, the first three terms is $|A|$ and the last three terms is the determinant of the matrix $\begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$.

Notice also that the first and second rows of this last matrix are identical hence by property (2) of determinants it must be zero. We therefore have

$$|B| = |A| + c.O = |A|$$

Thus adding a scalar multiple of a row (or column) to another does not change the determinant of the matrix.

5. If A is an $n \times n$ - matrix then $|A^t| = |A|$. This property is equally easy to see because if we use any row or column to evaluate the determinant, we obtain the same result. If we use the first rows to evaluate the determinant of A to obtain $|A|$. This first row in A becomes first column in A^t then use this first column to obtain $|A^t|$. By the above remark that determinant is independent of row or column used, the two values must be equal, hence $|A^t| = |A|$.

For a square matrix of large size (usually $n \times n$ matrix where $n > 3$) we now give a simple method of calculating the determinant. The method rely heavily on property 4 described earlier which states that if we obtain a matrix B from A by adding a scalar multiple of one row to another row then $|A| = |B|$. We start this procedure by subtracting (or adding) scalar multiples of one row of matrix A from all other rows of A so as to get a column of only one non-zero term. We then use this column to calculate the determinant.

We can continue this process until we reduce the determinant to a 2×2 case.

Example 5.2.1

Let $A = \begin{pmatrix} 3 & 4 & 6 & 1 \\ 4 & -2 & 2 & -1 \\ 1 & 3 & 1 & 1 \\ 3 & 1 & -5 & 1 \end{pmatrix}$, compute $|A|$.

Solution

A is a 4×4 matrix, to compute the determinant, we can make use of the first column, the least entry there is 1 and it is in the third row first column. We therefore row reduce this matrix until we get all entries in the first column to be zero except the entry in 3rd row first column which is 1. We apply the following row reduction.

$$R_1 \to -3R_3 + R_1$$

$$R_2 \to -4R_3 + R_2$$

$$R_4 \to -3R_3 + R_4$$

$$A = \begin{pmatrix} 3 & 4 & 6 & 1 \\ 4 & -2 & 2 & -1 \\ 1 & 3 & 1 & 1 \\ 3 & 1 & -5 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & -5 & 3 & -2 \\ 0 & -14 & -2 & -5 \\ 1 & 3 & 1 & 1 \\ 0 & -8 & -8 & -2 \end{pmatrix}$$

Since the type of row reduction used here preserves the determinant hence we use the first column to get the determinant.

$$\therefore |A| = \begin{vmatrix} 0 & -5 & 3 & -2 \\ 0 & -14 & -2 & -5 \\ 1 & 3 & 1 & 1 \\ 0 & -8 & -8 & -2 \end{vmatrix} = \begin{vmatrix} -5 & 3 & -2 \\ -14 & -2 & -5 \\ -8 & -8 & -2 \end{vmatrix}$$

We can row reduce further to reduce both entries in second row third column and third row third column to zero.

$$R_3 \rightarrow R_1 - R_3$$

$$R_2 \rightarrow \frac{-5}{2} R_3 + R_2$$

$$\begin{pmatrix} -5 & 3 & -2 \\ -14 & -2 & -5 \\ -8 & -8 & -2 \end{pmatrix} \rightarrow \begin{pmatrix} -5 & 3 & -2 \\ 6 & 18 & 0 \\ 3 & 11 & 0 \end{pmatrix}$$

We then get

$$|A| = \begin{vmatrix} -5 & 3 & -2 \\ 6 & 18 & 0 \\ 3 & 11 & 0 \end{vmatrix} = -2(66 - 54) = -24$$

Definition 5.2.1

An n × n matrix A is called non-singular if $|A| \neq 0$ and it is called singular if $|A| = 0$.

5.3 Application of Determinants

Let A be a non-singular n × n matrix, where $A = (a_{ij})$.

Recall that the minor of a_{ij} denoted by M_{ij} is the determinant of the $(n - 1) \times (n - 1)$ left when we delete the ith row and the jth column having a_{ij} as its entry. The

$$a_{ij} \text{ co-factor} = A_{ij} = (-1)^{i+j} M_{ij}$$

If

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdot & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & \cdot & \cdot & a_{nn} \end{pmatrix},$$

when we calculate all the cofactors then we get the matrix of co-factors as follows

$$\begin{pmatrix} A_{11} & A_{12} & \cdot & \cdot & \cdot & A_{1n} \\ A_{21} & A_{22} & \cdot & \cdot & \cdot & A_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ A_{n1} & A_{n2} & \cdot & \cdot & \cdot & A_{nn} \end{pmatrix}$$

The transpose of the above matrix of co-factors of A is called the Adjoint of A and is denoted by Adj (A).

$$\therefore \text{Adj}(A) = \begin{pmatrix} A_{11} & A_{21} & \cdot & \cdot & \cdot & A_{n1} \\ A_{12} & A_{22} & \cdot & \cdot & \cdot & A_{n2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ A_{1n} & A_{2n} & \cdot & \cdot & \cdot & A_{nn} \end{pmatrix}$$

Example 5.3.1

Let $A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 3 & 0 \\ 2 & 2 & 5 \end{pmatrix}$, compute Adj(A).

Solution

We first compute the matrix of co-factors and then take the transpose to get Adj(A).

$$A_{11} = \begin{vmatrix} 3 & 0 \\ 2 & 5 \end{vmatrix} = 15 - 0 = 15$$

$$A_{12} = -\begin{vmatrix} 2 & 0 \\ 2 & 5 \end{vmatrix} = -(10 - 0 = -10$$

$$A_{13} = \begin{vmatrix} 2 & 3 \\ 2 & 2 \end{vmatrix} = (4 - 6) = -2$$

$$A_{21} = -\begin{vmatrix} -1 & 0 \\ 2 & 5 \end{vmatrix} = -(-5 - 0) = +5$$

$$A_{22} = \begin{vmatrix} 1 & 0 \\ 2 & 5 \end{vmatrix} = (5 - 0) = 5$$

$$A_{23} = -\begin{vmatrix} 1 & -1 \\ 2 & 2 \end{vmatrix} = -(2 - (-2)) = -4$$

$$A_{31} = -\begin{vmatrix} -1 & 0 \\ 3 & 0 \end{vmatrix} = -(0 - (0)) = 0$$

$$A_{32} = -\begin{vmatrix} 1 & 0 \\ 2 & 0 \end{vmatrix} = -(0-0) = 0$$

$$A_{33} = \begin{vmatrix} 1 & -1 \\ 2 & 3 \end{vmatrix} (3-(-2)) = 5$$

$$\text{Matrix of co-factors of } A = \begin{pmatrix} 15 & -10 & -2 \\ 5 & 5 & -4 \\ 0 & 0 & 5 \end{pmatrix}$$

$$\text{Adj } A = \begin{pmatrix} 15 & 5 & 0 \\ -10 & 5 & 0 \\ -2 & -4 & 5 \end{pmatrix}$$

Example 5.3.2

Let $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$

(i) $|A| = a_{i1}A_{i1} - a_{i2}A_{i2} + a_{i3}A_{i3}, \quad i = 1,2,3 = \sum_{i=1}^{3} a_{ii}A_{ii}$

(ii) $a_{i1}A_{j1} + a_{i1}A_{j2} + a_{i3}A_{j3} = 0$ for $i \neq j$.

This means that the linear combination of the co-factors of a row with coefficients from another row is always zero.

Solution

(i) is clear because it is the expansion of $|A|$ using the ith row.

(ii) Suppose for instance i=2 and j=1, i.e. suppose we are taking the linear combination of the co-factors in the first row of the matrix of co-factors with coefficients in the second row of matrix A, i.e.

$$a_{21}A_{11} - a_{22}A_{12} + a_{23}A_{13} = a_{21}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{22}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{23}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$= \begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

Since the first and second rows of this matrix are equal then the determinant is zero, hence

$$a_{21}A_{11} + a_{22}A_{12} + a_{23}.A_{13} = 0.$$

The above result can be generalised to $n \times n$ matrix. If $A = (a_{ij})$ is an $n \times n$ matrix then

$$|A| = a_{i1}A_{i1} + a_{i2}A_{i2} + \ldots + a_{in}A_{in}$$

but if $i \neq k$ then

$$a_{i1}A_{k1} + a_{i2}A_{k2} + \ldots a_{in}A_{kn} = 0$$

This important result is applied to prove the following theorem which is one of important applications of determinants.

Theorem 5.3.1

If A is non-singular n × n matrix then A (adj (A)) = |A|.I

Proof:

$$\text{Let } A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix} \text{ then, } \operatorname{Adj}(A) = \begin{pmatrix} A_{11}\ A_{21}, \cdots, A_{n1} \\ A_{12}\ A_{22}, \cdots A_{n2} \\ \cdots \\ \cdots \\ \cdots \\ A_{1n}\ A_{2n}, \cdots A_{nn} \end{pmatrix}$$

hence

$$A.(\operatorname{Adj}(A)) = \begin{pmatrix} a_{11}\, a_{12} \cdots a_{1n} \\ a_{21}\, a_{22} \cdots a_{2n} \\ \cdots \\ \cdots \\ \cdots \\ a_{n1}\, a_{n2} \cdots a_{nn} \end{pmatrix} \begin{pmatrix} A_{1n}\, A_{21} \cdots A_{n1} \\ A_{12}\, A_{22} \cdots A_{n2} \\ \cdots \\ \cdots \\ \cdots \\ A_{1n}\, A_{2n} \cdots A_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^{n} a_{1j} A_{1j} & \sum_{j=1}^{n} a_{1j} A_{2j} & \cdots & \sum_{j=1}^{n} a_{1j} A_{nj} \\ \sum_{j=1}^{n} a_{2j} A_{1j} & \sum_{j=1}^{n} a_{2j} . A_{2j} & \cdots & \sum_{j=1}^{n} a_{2j} A_{nj} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \sum_{j=1}^{n} a_{nj} A_{1j} & \sum_{j=1}^{n} a_{nj} A_{2j} & \cdots & \sum_{j=1}^{n} a_{nj} A_{nj} \end{pmatrix}$$

From the generalisation of example 5.3.2 we see that $\sum a_{ij} A_{kj} = 0$ if $i \neq k$ and $\sum a_{ij} A_{kj} = |A|$ if $i = k$. It then follows that all the diagonal entries in the above matrix equals |A| while all other entries equals 0. The above matrix therefore reduces to

$$\begin{pmatrix} |A| & 0 & \cdots & 0 & 0 \\ 0 & |A| & \cdots & 0 & 0 \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 0 & 0 & \cdots & 0 & |A| \end{pmatrix} = |A|.I$$

hence

$$A(\mathrm{adj}(A)) = |A|.I.$$

This proves the theorem.

Since A is non-singular then $|A| \neq 0$ and so we can divide both sides of this equation by $|A|$ to have

$$\frac{A.(\mathrm{adj}(A))}{|A|} = I.$$

This will give us the equation

$$\frac{\mathrm{adj}(A)}{|A|} = A^{-1}.$$

This is a very important equation which we can use to compute the inverse of a non-singular matrix. When we need to find the inverse of a matrix, we first find the adjoint then we calculate the determinant. The inverse is then the adjoint divided by the determinant. Also the above equation shows that a square matrix is invertible if and only if its determinant is non-zero.

Example 5.3.3

Let $A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 3 & 0 \\ 2 & 2 & 5 \end{pmatrix}$ Find A^{-1} using adjoint method.

Solution

In example 5.3.1,.we computed the adjoint of A and got it to be

$$\mathrm{adj}(A) = \begin{pmatrix} 15 & 5 & 0 \\ -10 & 5 & 0 \\ -2 & -4 & 5 \end{pmatrix}$$

We then need to compute the determinant of A. It is easier to use the third column since only one entry is non-zero.

$$|A| = 5\begin{vmatrix} 1 & -1 \\ 2 & 3 \end{vmatrix} = 5(3-(-2)) = 25$$

hence

$$A^{-1} = \frac{\mathrm{adj}(A)}{|A|} = \frac{1}{25}\begin{pmatrix} 15 & 5 & 0 \\ -10 & 5 & 0 \\ -2 & -4 & 5 \end{pmatrix} = \begin{pmatrix} \frac{15}{25} & \frac{5}{25} & 0 \\ -\frac{10}{25} & \frac{5}{25} & 0 \\ -\frac{2}{25} & -\frac{4}{25} & \frac{5}{25} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{3}{5} & \frac{1}{5} & 0 \\ -\frac{2}{5} & \frac{1}{5} & 0 \\ \frac{2}{25} & -\frac{4}{25} & \frac{1}{5} \end{pmatrix}$$

When the size of a matrix is large and we need to compute the inverse, the adjoint method becomes cumbersome and inefficient. We will now describe a method of calculating inverses of non-singular matrices which does not involve using adjoint. This method is more efficient for large size matrices. Let A be the given n × n matrix, we pick any row with a non-zero first entry and add suitable scalar multiple of that row to other rows of A so as to reduce other entries, in the column, where that entry occurs, to zero. We then choose another row of this new matrix with a non-zero second entry and we perform similar operation to reduce other entries in that column to zero.

When we carry out this process n times then on each row we have only one non-zero entry and in each column also we have only one non-zero entry. We can then interchange the rows so as to get all the non-zero entries in the diagonal.

The resulting matrix is a diagonal matrix with say entries $a_1, a_2, \ldots a_n$ in the diagonal hence by property (4) of determinant then $|A| = \pm a_1 a_2, \ldots a_n$, since A is non-singular. We then multiply the first row by a_1^{-1}, second row by $a_2^{-1} \ldots$ nth row by a_n^{-1}.

If we perform the same operation on the identity matrix I, the final matrix we will obtain is exactly A^{-1}. To apply this procedure it is better and neater to write the matrix A and the identity matrix I side by side and perform the operation simultaneously.

Example 5.3.4

If $A = \begin{pmatrix} 1 2 2 \\ 3 1 0 \\ 1 1 1 \end{pmatrix}$, find A^{-1}.

Solution

We consider

$$(A : I) = \begin{pmatrix} 1 2 2 : 1 0 0 \\ 3 1 0 : 0 1 0 \\ 1 1 1 : 0 0 1 \end{pmatrix}$$

We perform the following operations on the above

$$R_2 \rightarrow -3R_1 + R_2$$

$$R_3 \rightarrow R_1 - R_3$$

then we get

$$\left(\begin{array}{ccccccc} 1 & 2 & 2 & : & 1 & 0 & 0 \\ 0 & -5 & -6 & : & -3 & 1 & 0 \\ 0 & 1 & 1 & : & 1 & 0 & -1 \end{array}\right)$$

We then perform the following operations on the above matrix.

$$R_1 \rightarrow \frac{5}{2} R_1 + R_2$$

$$R_2 \rightarrow 5 R_3 + R_2$$

and we get

$$\left(\begin{array}{ccccccc} \frac{5}{2} & 0 & -1 & : & \frac{-1}{2} & 1 & 0 \\ 0 & -5 & -6 & : & -3 & 1 & 0 \\ 0 & 0 & -1 & : & 2 & 1 & -5 \end{array}\right)$$

We further perform the following operation on the above.

$$R_1 \rightarrow R_1 - R_3$$

$$R_2 \rightarrow \frac{-1}{6} R_2 + R_3$$

to have

$$\left(\begin{array}{ccccccc} \frac{5}{2} & 0 & 0 & : & \frac{-5}{2} & 0 & 5 \\ 0 & \frac{5}{6} & 0 & : & \frac{5}{2} & \frac{5}{6} & -5 \\ 0 & 0 & -1 & : & 2 & 1 & -5 \end{array}\right)$$

Lastly we perform the following operations.

$$R_1 \rightarrow \frac{2}{5} R_1$$

$$R_2 \rightarrow \frac{6}{5} R_2$$

$$R_3 \rightarrow -R_3$$

we finally have $\left(\begin{array}{ccccccc} 1 & 0 & 0 & : & -1 & 0 & 2 \\ 0 & 1 & 0 & : & 3 & 1 & -6 \\ 0 & 0 & 1 & : & -2 & -1 & 5 \end{array}\right)$

hence

$$A^{-1} = \begin{pmatrix} -1 & 0 & 2 \\ 3 & 1 & -6 \\ -2 & -1 & 5 \end{pmatrix}$$

5.4 Linear Equations

We wish to apply what we have learnt so far in this chapter to solving a system of n simultaneous linear equations. Consider the following system of linear equations in unknowns $(x_1, x_2, \dots, x_n)$

$$\begin{aligned} a_{11}x_1+a_{12}x_2+\dots+a_{1n}x_n &= b_1 \\ a_{21}x_1+a_{22}x_2+\dots+a_{2n}x_n &= b_2 \\ \vdots \qquad \vdots \qquad &\quad \vdots \\ a_{n1}x_1+a_{n2}x_2+\dots+a_{nn}x_n &= b_n \end{aligned}$$

This equation can be written in matrix form as

$$\begin{pmatrix} a_{11}\,a_{12}\cdots a_{1n} \\ a_{21}\,a_{22}\cdots a_{2n} \\ \vdots \quad \vdots \quad \vdots \\ \vdots \quad \vdots \quad \vdots \\ \vdots \quad \vdots \quad \vdots \\ a_{n1}\,a_{n2}\dots a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ \vdots \\ b_n \end{pmatrix}$$

We can further write the above as

$$AX=B$$

where $A=(a_{ij})$ = matrix of the coefficients of the unknowns, X the unknowns and B is the column vector

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ \vdots \\ b_n \end{pmatrix}$$

If $|A| \neq 0$ then the matrix of coefficients is non-singular then A^{-1} exists and $A^{-1}A=AA^{-1}=I$.

In the equation $AX=B$, if we multiply both sides by A^{-1}, we have

$$A^{-1}AX=A^{-1}B$$

then

$$X=A^{-1}B.$$

This shows that

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = A^{-1} \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{pmatrix}$$

This shows that if $|A| \neq 0$ then the system of simultaneous linear equation $AX = B$ has a unique solution given by.

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = A^{-1}B = A^{-1} \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{pmatrix}$$

Example 5.4.1

By using the inverse of the matrix of coefficients, solve the following system of linear equations.

$$3x - 5 = 4y + 6z$$
$$2y - 4z = 3 - x$$
$$2x + 2z = 7 + 5y$$

Solution

We can rearrange the equation in the usual form to have the following.

$$3x - 4y - 6z = 5$$
$$x + 2y - 4z = 3$$
$$2x - 5y + 2z = 7$$

In matrix form, the above equations is of the form.

$$\begin{pmatrix} 3 & -4 & -6 \\ 1 & 2 & -4 \\ 2 & -5 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ x \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \\ 7 \end{pmatrix}$$

The matrix of coefficient is

$$A = \begin{pmatrix} 3 & -4 & -6 \\ 1 & 2 & -4 \\ 2 & -5 & 2 \end{pmatrix}$$

$$|A| = \begin{vmatrix} 3 & -4 & -6 \\ 1 & 2 & -4 \\ 2 & -5 & 2 \end{vmatrix} = 46$$

hence $|A| \neq 0$ which implies the equation has a unique solution. We need to compute A^{-1}. We can use the adjoint method or the reduction method discussed earlier, to compute A^{-1}. We use the adjoint method, we first compute the co-factors of A.

$$A_{11} = -16,\ A_{12}, = -10,\ A_{13} = -9$$

$$A_{21} = 38,\ A_{22} = 18,\ A_{23} = 7$$

$$A_{31} = 28,\ A_{32} = 6,\ A_{33} = 10$$

$$\text{Adj}\,(A) = \begin{pmatrix} -16 & 38 & 28 \\ -10 & 18 & 6 \\ -9 & 7 & 10 \end{pmatrix}$$

$$A^{-1} = \frac{\text{Adj}(A)}{|A|} = \frac{1}{46}\begin{pmatrix} -16 & 38 & 28 \\ -10 & 18 & 9 \\ -9 & 7 & 10 \end{pmatrix} = \begin{pmatrix} \frac{-16}{46} & \frac{38}{46} & \frac{28}{46} \\ \frac{-10}{46} & \frac{18}{46} & \frac{6}{46} \\ \frac{-9}{46} & \frac{7}{46} & \frac{10}{46} \end{pmatrix}$$

The unique solution is

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \vdots \\ \cdot \\ x_n \end{pmatrix} = A^{-1}\begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{pmatrix}$$

hence

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{-16}{46} & \frac{38}{46} & \frac{28}{46} \\ \frac{-10}{46} & \frac{18}{46} & \frac{6}{46} \\ \frac{-9}{46} & \frac{7}{46} & \frac{10}{46} \end{pmatrix}\begin{pmatrix} 5 \\ 3 \\ 7 \end{pmatrix} = \begin{pmatrix} \frac{-80}{46} + \frac{114}{46} + \frac{196}{46} \\ \frac{-50}{46} + \frac{54}{46} + \frac{42}{46} \\ \frac{-45}{46} + \frac{21}{46} + \frac{70}{46} \end{pmatrix} = = \begin{pmatrix} 5 \\ 1 \\ 1 \end{pmatrix}$$

hence

$$x = 5,\ y = 1 \text{ and } z = 1.$$

If we however have a system of homogeneous linear equations, 110313091

$$
\begin{aligned}
a_1x_1+a_{12}x_2+\dots+a_{1n}x_n&=0\\
a_{21}x_1+a_{22}x_2+\dots+a_{2n}x_n&=0\\
&\vdots\\
a_{n1}x_1+a_n2x_2+\dots+a_{nn}x_n&=0
\end{aligned}
$$

which we can write in matrix form as

$$
\begin{pmatrix}
a_{11} & a_{12} & \dots & a_{1n}\\
a_{21} & a_{22} & \dots & a_{2n}\\
\cdot & \cdot & & \cdot\\
\cdot & \cdot & & \cdot\\
\cdot & \cdot & & \cdot\\
a_n1 & a_n2 & \dots & a_{nn}
\end{pmatrix}
\begin{bmatrix}
x_1\\ x_2\\ \cdot\\ \cdot\\ \cdot\\ x_n
\end{bmatrix}
=
\begin{pmatrix}
0\\ 0\\ \cdot\\ \cdot\\ \cdot\\ 0
\end{pmatrix}
$$

If A is the above matrix of the coefficients, if $|A| \neq 0$ then A^{-1} exists and as in the previous result,

$$A^{-1}AX=A^{-1}.0=0$$

hence

$$
X=\begin{bmatrix}
x_1\\ x_2\\ \cdot\\ \cdot\\ \cdot\\ x_n
\end{bmatrix}
=
\begin{pmatrix}
0\\ 0\\ \cdot\\ \cdot\\ \cdot\\ 0
\end{pmatrix}
$$

This shows that the only possible solution of the system of homogeneous linear equations, when the matrix A of coefficients is non-singular, are the trivial solution.

$$x_1=0,\ x_2=0,\dots,x_n=0.$$

We lastly discuss another method of solving systems of linear equations without computing the inverse of the matrix of coefficients. In this method we only need to compute determinants of non-singular matrices. This method is known as *Cramer's Rule.*

Theorem 5.4.1

If

$$
\begin{aligned}
a_{11}x_1+a_{12}x_2+a_{1n}x_n&=b_1\\
a_{21}x_1+a_{21}x_2+a_{2n}x_n&=b_2\\
&\vdots\\
a_{n1}x_1+a_{n1}x_2+a_{nn}x_n&=b_n
\end{aligned}
$$

is a system of linear equations. Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

be the non-singular matrix of coefficients of the unknown $(x_1, x_2, \ldots, x_n)$. If B_i is the $n \times n$ matrix obtained from A by replacing the ith column of A by

$$B = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{pmatrix}$$

where $i = 1, 2, \ldots, n$, then the above equation has a unique solution given by

$$x_1 = \frac{|B_1|}{|A|}, x_2 = \frac{|B_2|}{|A|}, \ldots, x_n = \frac{|B_n|}{|A|}$$

Proof:

In the proof of this theorem, we will make use of the formula $A^{-1} = \frac{\mathrm{Adj}(A)}{|A|}$ which we have earlier derived. Since A is non-singular, A^{-1} exists and so

$$X = A^{-1}B = \left(\frac{1}{|A|}\mathrm{Adj}(A)\right)B.$$

On substituting for the above values, we have

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = \frac{1}{|A|}\begin{bmatrix} A_{11}\,A_{21}\cdots A_{n1} \\ A_{12}\,A_{22}\cdots A_{n2} \\ \cdot\ \cdot\ \cdot \\ \cdot\ \cdot\ \cdot \\ \cdot\ \cdot\ \cdot \\ A_{1n}\,A_{2n}\cdots A_{nn} \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{|A|} \begin{bmatrix} b_1A_{11}+b_2A_{21}+\dots+b_nA_{n1} \\ b_1A_{12}+b_2A_{22}+\dots+b_nA_{n2} \\ \cdot \\ \cdot \\ \cdot \\ b_1A_{1n}+b_2A_{2n}+\dots+b_nA_{nn} \end{bmatrix}$$

then equating the corresponding terms, we have

$$x_1 = \frac{b_1A_{11}+b_2A_{21}+\dots+b_nA_{n1}}{|A|}$$

$$x_2 = \frac{b_1A_{12}+b_2A_{21}+\dots+b_nA_{n2}}{|A|}$$

$$\vdots$$

$$x_n = \frac{b_1A_{1n}+b_2A_{2n}+\dots+b_nA_{nn}}{|A|}$$

If we replace the first column of A by

$$B = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{pmatrix}$$

we obtain

$$B_1 = \begin{pmatrix} b_1\, a_{12}\cdots a_{1n} \\ b_2\, a_{22}\cdots a_{2n} \\ \vdots \\ b_n\, a_{n2}\cdots a_{nn} \end{pmatrix}$$

and

$$|B_i| = b_1A_1 + b_2A_{21} + \dots + b_nA_{n1}$$

and in general if we replace the ith column of A by B we obtain the matrix

$$B_i \text{ and } |B_i| = b_1A_{1i} + b_2A_{2i} \dots + b_nA_i$$

hence

$$X_1 = \frac{b_1A_{11}+b_2A_{21}+\dots+b_nA_{n2}}{|A|} = \frac{|B_1|}{|A|}.$$

$$X_2 = \frac{b_1 A_{12} + b_2 A_{22} \dots + b_2 A_{n2}}{|A|} = \frac{|B_2|}{|A|}$$

$$X_i = \frac{b_1 A_{1i} + b_2 A_{22} + \dots + b_n A_{n2}}{|A|} = \frac{|B_i|}{|A|}$$

$$X_n = \frac{b_1 A_{1n} + b_2 A_{2n} + \dots + b_n A_{nn}}{|A|} = \frac{|B_n|}{|A|}$$

hence

$$X_1 = \frac{|B_1|}{|A|}, X_2 = \frac{|B_2|}{|A|}, \dots, X_n = \frac{|B_n|}{|A|}$$

Example 5.4.2
Solve the following system of linear equations using Crammers Rule.

$$x + y - z = 2$$
$$2x + z = 3$$
$$x - y + 4z = -1$$

Solution

The system of equation can be written in the form

$$\begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & 1 \\ 1 & -1 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix},$$

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & 1 \\ 1 & -1 & 4 \end{pmatrix}, X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, B = \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix}$$

$$B_1 = \begin{pmatrix} 2 & 1 & -1 \\ 3 & 0 & 1 \\ -1 & -1 & 4 \end{pmatrix}, B_2 = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 3 & 1 \\ 1 & -1 & 4 \end{pmatrix}, B_3 = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 0 & 3 \\ 1 & -1 & -1 \end{pmatrix}$$

$$|A| = \begin{vmatrix} 1 & 1 & -1 \\ 2 & 0 & 1 \\ 1 & -1 & 4 \end{vmatrix} = -4$$

$$|B_1| = \begin{vmatrix} 2 & 1 & -1 \\ 3 & 0 & 1 \\ -1 & -1 & 4 \end{vmatrix} = -8$$

$$|B_2| = \begin{vmatrix} 1 & 2 & -1 \\ 2 & 3 & 1 \\ 1 & -1 & 4 \end{vmatrix} = 4$$

$$|B_3| = \begin{vmatrix} 1 & 1 & 2 \\ 2 & 0 & 3 \\ 1 & -1 & -1 \end{vmatrix} = 4$$

$$x = \frac{|B_1|}{|A|} = \frac{-8}{-4} = 2$$

$$y = \frac{|B_2|}{|A|} = \frac{+4}{-4} = -1$$

$$z = \frac{|B_3|}{|A|} = \frac{+4}{-4} = -1$$

We cannot conveniently use crammers Rule or inverse of a matrix method always for solving system of linear equations. In fact if the variables are many then the computation of the determinant involved is very tedious. We will briefly discuss the elimination method which can in fact be used to solve any number of equations in any number of unknowns. Suppose we have m systems of linear equations in n unknowns as given below.

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m$$

For convenience, we start with the first equation where x_1 occurs and we add a suitable scalar multiple of other $(m-1)$ equations to it so as to eliminate the x_1 term in these other equations. We then have a new set of equations with exactly the same solution as the original set, but x_1 occurs only in the first equation. We now proceed to the second equation where x_2 term occur, we add a suitable scalar multiple of other $(m-2)$ equations to it so as to eliminate the x_2 term in these other equations.

We continue this procedure. When this procedure continues, it may happen that some of the new equations which arise, as a result of our elimination procedure, may have only zeros on the left hand side, if the corresponding constant on the right hand side is zero, we just disregard that equation because we have $0 = 0$. However, if the corresponding constant term is not zero then the original system of linear equations is inconsistent and therefore has no solution. When the process of elimination is completed then we start solving the equations from the last equation and continue to substitute upwards. This means that we use the last equation to solve for one of the variables in terms of the others and substitute in the next equation to the last one and so on.

Example 5.4.3

Solve the following system of linear equations.

$$3x + 3y - 3z = 0$$

$$-3x + 9y - 3z = 0$$

$$-6x + 6y = 0.$$

Solution

We want to eliminate x in equations 2 and 3. We add equations (1) and (2) and replace equation (2) by the result. Similarly we add 2 times equation (1) to equation 3, we then obtain the following.

$$\begin{aligned} 3x + 3y - 3z &= 0 \\ 0 + 12y - 6z &= 0 \\ 0 + 12y - 6z &= 0 \end{aligned}$$

To eliminate y in the third equation we subtract equation (3) from equation (2). We now have

$$\begin{aligned} 3x + 3y - 3z &= 0 \\ 0 + 12y - 6z &= 0 \\ 0 &= 0. \end{aligned}$$

The last equation is $0 = 0$ which can be ignored. We can further reduce the equations to a better form by multiplying equation (1) by $\frac{1}{3}$ and equation (2) by $\frac{1}{6}$. We then have

$$\begin{aligned} x + y - z &= 0 \\ 2y - z &= 0 \end{aligned}$$

On solving this equation we have

$$z = 2y.$$

Substituting this in the first equation we have

$$\begin{aligned} x = z - y = 2y - y &= y. \\ \therefore x &= y \\ z &= 2y, \end{aligned}$$

y is arbitrary. Notice that this system of linear equation has no unique solution in fact it has infinitely many solutions.

In the system of m linear equations in n unknowns, $x_1, x_2, \ldots x_n$, suppose the system is consistent. When we complete the system of elimination, we will still have m equations although some of them may be of the form $0 = 0$.

Suppose there are $(m - r)$ equations of the form $0 = 0$. This means that we have r equations in which the variables appear. Each of this r equations can be used to solve for one of the variables say X_j for example in terms of the others $x_{j+1}, \ldots, x_n$.

It follows that in the final solution exactly r of the variables out of $x_1, \ldots x_n$ will be expressed in terms of the remaining $(n - r)$ variables. These $(n - r)$ variables will take arbitrary values. We say that r is the rank of the matrix of coefficient $A = (a_{ij})$ which is an $m \times n$ matrix and the variables are called the free variables. For instance in Example 5.4.3, there is only one free variable and the rank of the matrix is 2.

5.5 Exercise

✓ 1. If $A = \begin{pmatrix} 1 & a & a^2 & 0 \\ 0 & 1 & a & a^2 \\ a^2 & 0 & 1 & a \\ a & a^2 & 0 & 1 \end{pmatrix}$, show that $|A| = 1 + a^4 + a^8$

2. Show that if A and B are n × n matrices such that A is similar to B then $|A| = |B|$ hence or otherwise, if T: $R^3 \rightarrow R^3$ is defined by

$$T(x, y, z) = (3x - 2z, 5y + 7z, x + y + z)$$

Find det(T).

3. If $A = \begin{pmatrix} t-4 & 0 & 0 \\ -2 & t+1 & 1 \\ 3 & 4 & t-2 \end{pmatrix}$, find the value of t for which A is singular. ✓

4. An n × n matrix A is said to be orthogonal if $A'A = I$. Show that if A is an orthogonal matrix then $|A| = \pm 1$.

5. Find the inverse of matrix $A = \begin{pmatrix} 1 & 2 & 2 & 3 \\ 3 & -1 & 1 & -2 \\ 1 & 0 & -2 & 0 \\ 4 & -3 & 0 & 2 \end{pmatrix}$ ✓

6. If A and B are invertible n × n matrices, prove that $(AB)^{-1} = B^{-1}A^{-1}$ hence show that $(A^2)^{-1} = (A^{-1})^2$ ✓

7. If $A = \begin{pmatrix} 2 & -5 & 2 \\ 1 & 2 & -4 \\ 3 & -4 & -6 \end{pmatrix}$, compute A^{-1}. Use this result to solve the following system of linear equations.

$$2x + 2z = 7 + 5y$$
$$x + 2y - 3 = 4z$$
$$x - 6z = 4y + 5.$$ ✓

8. Use Crammer's Rule to solve the following equations.

$$3x + y + z = 1$$
$$6x - y = 7$$
$$6y + 2 = 5$$ ✓

9. Find a general solution for the following system of linear equations.

$$x + 2y - z - 3w + u = 4$$
$$2x + 3y + z + 2u = 10$$
$$y + 2w + u = 5.$$

10. Solve the following system of linear equations.

$$x + y + z = 3$$
$$2x + z = 4$$
$$2y + z = 2.$$ ✓

-8+2

-8+15 64+5

-36+15 4:2

11. For any 3 × 3 matrix $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ show that $Ad_j(Ad_j A) = A$

12. If A is an n × n triangular matrix and B is an n × n diagonal matrix
    (i) show that $Ad_j(A)$ is a triangular matrix and $Ad_j(B)$ is a diagonal matrix.
    (ii) If $A = (a_{ij})$ is an $n \times n$ triangular matrix and $B = (b_{ii})$ is an $n \times n \times$ diagonal matrix, show that A is invertible if and only if a_{ii} ne 0 and B is invertible if and only if $b_n \neq 0$.

13. Let V be the vector space of polynomials in variable x.

If $D:V \longrightarrow V$ be the differential operator on V i.e. if $f(x) \varepsilon V$ then $D(f(x)) = \frac{df}{dx}$.

(i) show that D is a linear map.
(ii) Find a matrix of D in the basis $\{e^{3x}, e^{3x}, x^2 e^{3x}\}$.

14. If $T: R^3 \longrightarrow R^3$ is a linear operator defined by $T(x, y, z) = (5x + 7y, 3x - 2z, 3x + y + z)$ compute det(T).

15. If $A = \begin{pmatrix} a_1, & 0, & 0, & \dots & 0, \\ 0, & a_2, & 0, & \dots & 0, \\ & & \vdots & \dots & \\ 0, & 0, & 0, & \dots & a_n, \end{pmatrix}$ show that

$$A = \begin{pmatrix} a_1^{-1}, & 0, & 0, & \dots & 0, \\ 0, & a_2^{-1}, & 0, & \dots & 0, \\ & & \vdots & \dots & \\ 0, & 0, & 0, & \dots & a_n^{-1} \end{pmatrix}$$

CHAPTER SIX

EIGENVALUES AND EIGENVECTORS

Let T be a linear operator on a finite dimensional vector space V. Given any basis of the vector space V, we have discussed how to represent this linear operator by a matrix relative to this given basis. We want to find an ordered basis of V for which the matrix of T will be as simple as possible. One of the simplest matrix is the diagonal matrix.

$$D = \begin{pmatrix} a_{11} & 0 & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

If it is possible to find a basis $v = (v_1, v_2, \dots v_n)$ for V in which T can be represented by D, then questions about the operator T will have ready answers by a quick glance at D. For instance, the determinant of T will just be product of the a_{ii}'s. If $v = (v_1, v_2, \dots v_n)$ is a basis of V then $[T]_v = D$ if and only if $T(v_i) = a_i v_i \; i = 1, \dots, n$. The range of T will be the subspace of V generated by v_i's such that $a_i \underline{=} 0$.

The remaining v_is generate the null space. This shows that the range of T and null space of T can be easily determined. In fact if for some basis v, $[T]_v = D$, any question about T can be readily answered. The question that readily come to mind is, can each linear operator T be represented by a diagonal matrix on some ordered basis? If yes, then how can we find the basis. If not, what simplest form of matrix can we represent T? These types of questions, we shall try to answer in this chapter.

Definition 6.1.1

Let V be a vector space over a field K and T a linear operator on V. A scalar c ε K is called an eigenvalue of T if there exists a non zero vector v ε V for which T(v) = cv. There may be more than one vector v which satisfies the condition T(v) = cv. However, every vector which satisfies this relation is called eigenvector belonging to eigenvalue c.

Theorem 6.1.1

The set of all eigenvectors corresponding to eigenvalue c ε K of linear operator T on a finite dimensional vector space V is a subspace of V.

Proof:
Let V_c represent the set of all eigenvectors corresponding to eigenvalue $c \varepsilon K$. If $u, v \varepsilon V_c$, then

$$T(u) = cu \text{ and } T(v) = cv$$

and

$$T(u+v) = T(u) + T(v) = cu + cv = c(u+v),$$

hence $u + v \varepsilon V_c$. Also if $a \varepsilon K$ and $u \varepsilon V_c$, then

$$T(au) = aT(u) = acu = c(au)$$

hence $a u \varepsilon V_c$. This shows that V_c is a subspace of V. This subspace is called eigen space associated with c. Some textbooks call eigenvalues, as characteristic roots, proper values or special values. In this book we shall use "eigenvalues".

Theorem 6.1.2:
Let T be a linear operator on a finite dimensional vector space V over a field K. The scalar $c \varepsilon K$ is an eigenvalue of T if and only if the operator (T-cI) is singular. The eigenspace of c is the Kernel of (T - cI).

Proof
By definition, $c \varepsilon K$ is an eigen value of T if and only if there exists a non zero vector $v \varepsilon V$ such that $T(v) = cv$.
This implies

$$T(v) - cv = 0, (T- cI)(v) = 0.$$

For this to have a non zero solution v, det(T - cI) = 0 and this implies that (T - cI) must be singular.
Also $v \varepsilon V$ belongs to eigen space of $c \varepsilon K$ if and only if (T- cI)(v) = 0 hence $v \varepsilon$ Kernel of (T - cI).
This also shows that the eigen space of c is the null space of (T - cI).

The fact that $c \varepsilon K$ is an eigen value of T if and only if det (T - cI) = 0 is a very important result because it tells us where to look for eigenvalues of T. Notice that det(T - cI) is a polynomial of degree n in variable c and so we only need to find these roots and this will give us the eigenvalues of T. If $v = (v_1, v_2, \ldots v_n)$ is a basis for V and $[T]_v = A$ then (T - cI) is invertible if and only if (A - cI) is invertible. This leads us to the following.

Definition 6.1.2
Let A be an n x n matrix over a field K, an eigenvalue of A in K is a scalar $c \varepsilon K$ such that det (A - cI) = 0.

6.2 Characteristics Polynomial

Since c is an eigenvalue of A if and only if det (A - cI) = 0, we can then form the matrix (A - xI) with polynomial entries and let the polynomial f(x) = det(A - xI).

Therefore the eigenvalues of A are the roots c of the polynomial f(x). It follows that $c \varepsilon K$ is an eigenvalue of A if and only if $f(c) = 0$. The polynomial f(x) is called the characteristic polynomial of A. When A is an $n \times n$ matrix, then the characteristic polynomial f(x) of A is a monic polynomial of degree n.

Example 6.2.1

Let T be the linear operator on R^2 defined by $T(x,y) = (4x + y, 3x + 2y)$. Find the eigenvalue of T.

Solution

In standard ordered basis the operator T is represented by the matrix.

$$A = \begin{pmatrix} 4 & 1 \\ 3 & 2 \end{pmatrix}$$

The characteristic polynomial for A is

$$\det(A - xI) = \begin{vmatrix} 4-x & 1 \\ 3 & 2-x \end{vmatrix} = (4-x)(2-x) - 3 = 0$$

This implies $8 - 6x + x^2 - 3 = 0$

$$x^2 - 6x + 5 = 0$$
$$(x-5)(x-1) = 0$$
$$x = 5, x = 1$$

The eigenvalues of A are 5 and 1. To find the eigenvectors, when $x = 1$, we have

$$A - I = \begin{pmatrix} 3 & 1 \\ 3 & 1 \end{pmatrix}$$

from this matrix we see that A - I has rank 1 and the nullity is 1. The space of eigenvectors associated with $x = 1$ is one dimensional. Using the above, then see that if $v = (x,y)$ is an eigen vector, then

$$\begin{pmatrix} 3 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

hence

$$3x + y = 0$$
$$3x + y = 0$$
$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$$

This shows that $v = (-1, 3)$ spans the null space of A - I.

When $x = 5$ then $A - 5I = \begin{pmatrix} -1 & 1 \\ 3 & -3 \end{pmatrix}$

Also this matrix has rank 1 and the null space is also 1 hence the eigenspace associated with eigen value x=5 has dimension 1. From above the eigen space is generated by $(x,y) = (1,1)$, i.e. $T(v) = 5v$ if and only if v is a scalar multiple of (1, 1).

Example 6.2.2

Let T be a linear operator on R^3 defined by T(x,y,z)=(3x+y+z, 2x + 4y + 2z, x + y +3z)
Find the eigenvalues and the eigenvectors of T.

Solution

The matrix representing T on the usual basis is

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ 1 & 1 & 3 \end{pmatrix}$$

Characteristic polynomial of A is

$$f(x) = \det(A - xI) = \begin{vmatrix} 3-x & 1 & 1 \\ 2 & 4-x & 2 \\ 1 & 1 & 3-x \end{vmatrix}$$

$$= (3-x)\{(4-x)(3-x)-2\} - 1\{2(3-x)-2\} + 1\{2-(4-x)\}$$
$$= (3-x)\{(0-7x+x^2\} + (2x-4) + (x-2)$$
$$= (3-x)(x-2)(x-5) + 2(x-2) + (x-2)$$
$$= (x-2)\{(3-x)(x-5)+2+1\} = (x-2)(8x-12-x^2)$$
$$f(x) = (x-2)(x-2)(x-6) = (x-2)^2(x-6) = 0$$
$$\therefore x = 2 \text{ or } 6.$$

The eigenvalues are 2 and 6. We now find the eigenvectors for each eigenvalue. When $x = 2$ then

$$A - 2I = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{pmatrix}$$

By examining the above we see that the matrix has rank 1 hence the nullity of A - 2I is 2. This means that the eigenspace of 2 is generated by two independent vectors. The independent vectors can be obtained thus, if $v = (x, y, z)$ is an eigenvector corresponding to eigenvalue 2 then

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This implies

$$x + y + z = 0$$
$$2x + 2y + 2z = 0$$
$$x + y + z = 0$$

All the above equations imply

$$x + y + z = 0.$$

It follows that

$$z = -(x + y)$$

where x and y are arbitrary. If $x = 1$, $y = -1$ then

$$z = 0,$$

hence an eigenvector $v_1, = (1, -1, 0)$.

If $x = 1, y = 0$ then $z = -1$ the second eigenvector is given by

$$v_2 = (1, 0, -1).$$

These two vectors v_1, v_2 are independent and hence form a basis for the eigenspace. This means that any vector $v \varepsilon R^3$, $T(v) = 2v$ if and only if v is a linear combination of v_1 and v_2. Now when $x = 6$

$$A - 6I = \begin{bmatrix} -3 & 1 & 1 \\ 2 & -2 & 2 \\ 1 & 1 & -3 \end{bmatrix}$$

By examining this it may not be easy to see that the rank of this matrix is 2 but if we now reduce to echelon form, we will have a matrix of the form.

$$\begin{pmatrix} -3 & 1 & 1 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

This clearly shows that the above matrix has rank 2 and so the nullity of $T - 6I$ is 1. This means that the eigenspace of 6 is generated by one vector. To find the vector, if $v = (x, y, z)$ is in the eigen space, then

$$\begin{pmatrix} -3 & 1 & 1 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Or

$$-3x + y + z = 0$$

and

$$-y + 2z = 0$$

hence

$$y = 2z$$
$$3x = y + z = 3z$$

hence

$$x = z$$

where z is arbitrary, if $z = 1$, then

$$x = 1 \text{ and } y = 2.$$

A generator of the eigenspace of 6 is the vector $v = (1, 2, 1)$. The eigenvalues of the operators T and their eigenvectors are, eigenvalue = 2 with eigenvectors generated by $v_1 = (1, -1, 0)$ and $v_2 = (1, 0, -1)$, while another eigenvalue = 6 with eigenvector generated by $v = (1, 2, 1)$.

Example 6.2.3

Consider a linear operator T on R^2 defined by $T(x, y) = (y, -x)$. Find the eigenvalues of T.

Solution

In the usual basis, the matrix representing T is $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. The characteristic polynomial for T is

$$f(x) = \det(A - xI) = \begin{vmatrix} x & 1 \\ -1 & -x \end{vmatrix} = x^2 + 1.$$

This polynomial has no zero in the set R of real numbers and so T has no eigenvalue in R. If however, T is viewed as a linear operator on set C of complex numbers then T has eigenvalues i and -i. This example shows that a linear operator on a vector space may not have eigenvalue over a particular field but may have over another field.

Definition 6.2.2

Let T be a linear operator on a finite dimensional vector space V, we say that T is diagonalizable, if there exists a basis for V where each vector in the basis is an eigenvector for T. This definition makes sense because if $(v_1, v_2, \ldots, v_n)$ is a basis of V which are eigenvectors of T then

$$T(v_i) = c_i v_i$$

where i = 1, 2,..., n. The matrix of T in this basis will therefore be of the form

$$\begin{pmatrix} c_1 & 0 & 0 & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & c_n \end{pmatrix}$$

In example 6.2.3 we have a linear operator on R^2 which is not diagonalizable because it has no eigenvalues.

In Example 6.2.2 however, the operator T has a basis consisting of eigenvectors namely

$$v = (v_1 = (1,-1,0), \; v_2 = (1,0,-1) \text{ and } v_3 = (1,2,1)).$$

This shows that this operator is diagonalizable and the diagonal form is

$$D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

The fact that T is diagonalizable means that the original matrix A of T over the usual basis is similar to the diagonal matrix D. The matrix P which enable us to change coordinates from the basis v = (v_1, v_2, v_3) to the usual basis is the matrix whose column vectors are the eigenvectors. Hence

$$P = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 2 \\ 0 & -1 & 1 \end{bmatrix}$$

and so

$$P^{-1} A P = D$$

$$P^{-1} = \frac{1}{4}\begin{pmatrix} 2 & -2 & 2 \\ 1 & 1 & -3 \\ 1 & 1 & 1 \end{pmatrix}$$

$$P^{-1}AP = P^{-1}AP = \frac{1}{4}\begin{pmatrix} 2 & -2 & 2 \\ 1 & 1 & -3 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ 1 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 2 \\ 0 & -1 & 1 \end{pmatrix}$$

$$= \frac{1}{4}\begin{pmatrix} 4 & -4 & 4 \\ 2 & 2 & -6 \\ 6 & 6 & 6 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 2 \\ 1 & -1 & 1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 8 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 24 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{pmatrix} = D$$

We will now show in general that eigenvectors of a linear operator are linearly independent.

Theorem 6.2.1

Let T be a linear operator on a finite dimensional vector space V and let $(v_1, v_2, \ldots, v_n)$ be eigenvectors of T belonging to distinct eigenvalues $c_1, c_2, \ldots, c_n$. Then $(v_1, v_2, \ldots v_n)$ are linearly independent.

Proof:

The proof is by induction on n. If n = 1 and $a_1 v_1 = 0$, since $v_1 \neq 0$ then $a_1 = 0$. Assume the theorem is true for n - 1 eigenvectors and suppose $a_1 v_1 + a_2 v_2 + \ldots + a_n v_n = 0$ then we show that $a_i = 0$ for all i = 1,..., n. Applying T to the relation above, we have

$$T(a_1 v_1 + a_2 v_2 + \ldots + a_n v_n) = T(0) = 0$$

hence

$$a_1 T(v_1) + a_2 T(v_2) + \ldots + a_n T(v_n) = 0$$

Since v_i is an eigenvector for eigenvalue c_i then $T(v_i) = c_i v_i$. Substituting in the above, we have

$$a_1 c_1 v_1 + a_2 c_2 v_2 + \ldots + a_n c_n v_n = 0.$$

When we multiply the first equation by c_n, we have

$$a_1 v_1 c_n + a_2 v_2 c_n + \ldots + a_n v_1 c_n$$

And on subtracting it from the first equation, we have

$$a_1(c_1 - c_n)v_1 + \ldots + a_{n-1}(c_{n-1} - c_n)v_{n-1} = 0.$$

By induction hypothesis, $v_1, v_2, \ldots, v_n-1$ are linearly independent hence

$$a_1(c_1 - c_n) = 0, \quad a_2(c_2 - c_n) = 0, \ldots, a_n-1(c_n-1 - c_n) = 0.$$

However, each c_i is distinct i.e. $c_i \neq c_n$ for $i \neq n$ hence

$$a_1 = a_2 = \ldots = a_{n-1} = 0.$$

When we substitute this in the first equation, we have

$$a_n v_n = 0 \text{ and since } v_n \neq 0, \text{ then } a_n = 0.$$

This shows that $v_i, i = 1, 2, \ldots, n$ are linearly independent.

Theorem 6.2.2

Similar matrices have the same characteristic polynomial.

Proof

Let A be a matrix similar to B then $A = P^{-1}BP$ for a non-singular matrix P.

$$\det(A - xI) = \det(P^{-1}BP - xI)$$

but $I = P^{-1}IP$

$$\det(P^{-1}BP - xI) = \det(P^{-1}BP - xP^{-1}IP) = \det(P^{-1}(B - xI)P)$$

$$= \det(P^{-1})\det(B - xI)\det(P) = \det(B - xI)$$

This theorem shows us that the characteristic polynomial of a linear operator is the same for any $n \times n$ matrix which represents T in some ordered basis for the n-dimensional vector space V.

6.3 Minimal Polynomial

Let V be a vector space over a field K and K[x] the set of all polynomials with coefficient in K. If T is a linear operator on V and p(x) ε K [x], then p(T) is again a linear operator. For instance if

$$p(x) = \sum_{i=1}^{n} a_i x^i$$

where $a_i \in K$, then

$$p(T) = \sum_{i=1}^{n} a_i T^i.$$

Each term $a_i T^i$ is a linear operator because T^i is just the composite of T i times, hence p(T) is itself a linear operator. A polynomial p(x) ε K{x} is said to annihilate T if p(T)=0. Let A_T represent the collection of polynomials which annihilate T, if p(x), q(x) ε A_T then

$$p(T) = 0 \text{ and } q(T) = 0.$$

multh

It follows that

$$(p+q)(T) = p(T) + q(T) = 0 + 0 = 0$$

and

$$(gp)(T) = g(T)p(T) = 0.0 = 0$$

therefore

$$p+q,\ pq\ \varepsilon\ A_T$$

This shows that A_T is an ideal of K [x]. We need to show that this ideal A_T is not a zero ideal and so it contains a non-zero polynomial. Let T be a linear operator on an n-dimensional vector space V, if $L(V,V)$ represents the space of linear operators on V, $L(V,V)$ has dimension n^2. We will now examine the first n^2+1 linear operators which are powers of T. $T^o = I, T, T^2, T^3, \ldots T^{n^2}$.. Since each of these linear operators belong to $L(V,V)$ and $L(V,V)$ has dimension n^2 then these n^2+1 vectors must be linearly dependent hence there exists $a_1, a_2, \ldots a_{n^2}$ which are scalars in K such that $a_o I + a_1 T + a_2 T^2 + \ldots + a_{n2} T^{n^2} = 0$. If f(x) = $\sum_{i=1}^{n^2} a_i x_i$ then

$$f(T) = a_o I + a_1 T + \ldots + a_{n2} T^{n2} = 0$$

hence f(x) ε K[x] of degree n^2 and f(T) = 0. This implies f(x) ε A_T and so A_T is a non-zero ideal of K[x]. Recall that K[x] is a principal ideal domain and so every ideal in it must be principal, hence A_T is a principal ideal. This shows that the ideal A_T of annihilators of T is generated by a monic polynomial which we denote by p. It then follows that if f(x) ε A_T then there exists g(x) ε A_T such that f(x) = p(x) g(x).

Definition 6.3.1

Let V be a finite dimensional vector space over a field K and T a linear operator on V. The minimal polynomial for T is the unique monic generator of the ideal of polynomials over K which annihilates T. The name "minimal polynomial" comes from the fact that the generator of the ideal is the monic polynomial of least degree. It is characterised by the following properties.

(i) P is a monic polynomial over the scalar field K.
(ii) P(T) = 0
(iii) No polynomial over K in A_T has a smaller degree than p.

Let T be a linear operator on a vector space V and if A is an $n \times n$ matrix representing T in an ordered basis then f(T) is represented in the basis by f(A) hence f(T) = 0 if and only if f(A) = 0.

Theorem 6.3.1

Similar matrices have the same minimal polynomial.

Proof:

If A and B are similar, then $A = p^{-1} B P$ for some non-singular matrix P then

$f(A) = f(p^{-1} B P) = p^{-1} f(B) P$. This shows that f(A) = 0 if and only if f(B) = 0, hence similar matrices have the same minimal polynomial.

Theorem 6.3.2
Let T be a linear operator on an n-dimensional vector space V then the characteristic and minimal polynomial have the same roots except for multiples.

Proof:
Since the roots of the characteristic polynomial of T are the eigenvalues of the operator then it is enough to show that if p is the minimal polynomial for T and $c \varepsilon K$ then $p(c)=0$ if and only if c is an eigenvalue for T. Suppose $p(c)=0$ then it follows that $x-c$ divides $p(x)$, i.e.

$$p(x)=(x-c)q(x)$$

where $q(x) \varepsilon K[x]$ and $\deg(q(x)) < \deg(p(x))$.

Since p is the minimal polynomial then no other polynomial of lower degree annihilates T, in particular $q(T) \neq 0$,

$$p(T)=(T-cI)q(T)=0.$$

Since $q(T) \neq 0$ then $(T-cI)=0$ which means that c is an eigenvalue of the linear operator T.

Conversely suppose c is an eigenvalue of the linear operator T, then $T(v)=cv$ with $v \varepsilon V$ and $v \neq 0$. It then follows that $p(T)\,v=p(c)\,v$, since $p(x)$ is the minimal polynomial for T, then $p(T)=0$ and so $p(c)\,v=0$ and since $v \neq 0$ we have $p(c)=0$. This proves that the eigenvalue c is a root of $p(x)$.

The preceding theorem is a very useful one because it narrows down the search for the minimum polynomial for an operator. Given a linear operator T on an n-dimensional vector space V, with respect to an ordered basis we can compute the matrix A representing the linear operator. We can then find the characteristic polynomial of the matrix. Since the characteristic polynomial and minimal polynomial have the same roots. If the roots of the characteristic polynomial are $c_1, c_2, \ldots c_k$, the preceding theorem suggest to us that $p(x)$ will at least be of the form $(x-c_1)(x-c_2)\ldots(x-c_k)$.

To be able to determine precisely the minimal polynomial of a linear operator we will need the following theorem.

Theorem 6.3.3: (Cayley-Hamilton)
Every linear operator T is a zero of its characteristic polynomial.

Proof:

Let V be an n-dimensional vector space over a field K. Let $v=(v_1, v_2, \ldots v_n)$ be a basis for V and T a linear operator on V. Let A be the matrix which represents T in the given basis. The characteristic polynomial of T is

$$f(x)=|A-xI|=a_0+a_1x+\ldots+a_{n-1}x^{n-1}+x^n$$

Notice that $(A - xI)$ is an $n \times n$ matrix and let $(A - xI)_{ij}$ be the i,j co-factor of the matrix. Each of these co-factors is a polynomial of degree at most n-1. Let the classical adjoint be denoted by B(x) hence

$$B(x) = B_{n-1}x^{n-1} + \ldots + B_1 x + B_0$$

where each B_i is an $n \times n$ matrix over K, which is independent of x.

Recall that for any square matrix we have the formula

$$A.(\text{adj } A) = |A|.$$

Using this formula for our matrix A - xI, we have

$$(A - xI)B(x) = |A - xI|.I.$$

Substituting for B(x), we have

$$(A - xI)(B_{n-1}x^{n-1} + \ldots + B_1 x + B_0) = (x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0)I$$

opening the brackets, we have

$$B_{n-1}x^n + (B_{n-2} - AB_{n-1})x^{n-1} + \ldots + -AB_0 = x^nI + a_{n-1}x^{n-1}I + \ldots + a_1 xI + a_0 I$$

Equating coefficients of x^i, we have

$B_{n-1} = I$

$B_{n-2} - AB_{n-1} = a_{n-1}I$

$B_{n-3} - AB_{n-2} = a_{n-2}I$

$B_0 - AB_1 = a_1 I$

$-BA_0 = a_0 I$

multiplying the above equations by $A^n, A^{n-1}, \ldots A, I$ respectively we have

$$A^nB_{n-1} = A^n$$

$$A^{n-1}B_{n-2} - A^nB_{n-1} = a_{n-1}A^{n-1}$$

$$A^{n-2}B_{n-3} - A^{n-1}B_{n-2} = a_{n-2}A^{n-2}$$

We then add all the left hand sides of the equations together and this addition will give us zero. We also add all the terms in the right hand sides of the equations together and equate it to the sum of the left hand side which is zero. We then have.

$$0 = A^n + a_{n-1}A^{n-1} + a_{n-2}A^{2-2} + \ldots + a_1 A + a_0 I = f(A)$$

This shows that f(A) = 0 and so A is a zero of its characteristics polynomial.
At this stage, theorems 6.3.2 and 6.3.3 show us that for a linear operator T, the minimal polynomial p(x) divides the characteristics polynomial. This is because if f(x) is the characteristic polynomial then f(T) = 0 hence $f(x) \varepsilon A_T$ and so p(x) must divide f(x). Also the two polynomials f(x) and p(x) have the same roots. If

$$f(x) = c(x - c_1)^{d_1}(x - c_2)^{d_2}\ldots(x - c_k)^{d_k},$$

where $c_1, c_2, \ldots c_k$ are distinct and $d_i \geq 1$ then

$$p(x) = (x-c_1)^{r_1}(x-c_2)^{r_2}\ldots(x-c_k)^{r_k}, \quad 1 \leq r_j \leq d_j.$$

Suppose T is a linear operator on an n-dimensional vector space V and T is diagonalizable.

Let $c_1, c_2, \ldots c_n$ be distinct eigenvalues then it follows that the characteristic polynomial is a product of linear factors. Since the characteristic and minimal polynomial have the same roots then the minimal polynomial is also the product of linear factors, hence $p(x) = (x-c_1)(x-c_2)..(x-c_n)$. This shows that if an operator is diagonalizable then the minimal polynomial is a product of distinct linear factors.

Example 6.3.1

Let T be a linear operator on R^3 defined by T(x, y, z) = (2x, 2y + 2z, z). Find the minimum polynomial for T.

Solution

In the standard basis for R^3 the matrix of T is

$$A = \begin{pmatrix} 2, 0, 0 \\ 0\ 2\ 2 \\ 0\ 0\ 1 \end{pmatrix}$$

The characteristic polynomial of T is

$$\det(A - xI) = \begin{vmatrix} 2-x & 0 & 0 \\ 0 & 2-x & 2 \\ 0 & 0 & 1-x \end{vmatrix} = (2-x)(2-x)(1-x)$$

or

$$(x-2)^2(x-1) = 0$$

Since the minimum polynomial p(x) of T and the characteristic polynomial f(x) have the same roots and p(x) divides f(x) then p(x) can only be either of the following two forms $(x-2)(x-1)$ or $(x-2)^2(x-1)$.

Recall however that p(x) is the polynomial with smallest degree such that $P(A) = 0$.

Notice that the polynomial f(x) $(x-2)^2(x-1)$ is the characteristic polynomial of A and so $f(A) = 0$. We therefore only need to check if A is a zero of the polynomial $(x-2)(x-1)$.

$$(A-2I)(A-I) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0\,0\,0 \\ 0\,0\,0 \\ 0\,0\,0 \end{pmatrix}$$

hence A is a zero of $(x-2)(x-1)$ and so the minimum polynomial of A is $(x-2)(x-1)$.

6.4 Exercise

1. Let T be a linear operator on R^3 defined by T(x,y,z) = (x+2x+2y, x+2y-z, -x+y+4z). Find the eigenvalues and corresponding eigenvectors. Show that T is diagonalisable and write the diagonal form.

2. Find the minimum polynomial of the matrix $A = \begin{pmatrix} 1 & 4 & -3 \\ 0 & 3 & 1 \\ 0 & 2 & -1 \end{pmatrix}$.

3. Consider a linear operator T defined on K^3 where K is field, by T(x, y, z) = (3x, 2y - 5z, y - 2z). Show that T is diagonalizable if K = C the field of complex numbers but T is not diagonalisable if K = R = field of real numbers.

4. If A is an n × n matrix, show that A and A^t have the same eigenvalues.

5. Show that the matrix $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ is not diagonalizable and give reasons.

6. Let T be a linear operator on a vector space V. If u ε V is an eigen vector of T corresponding to eigenvalue c show that u is also an eigenvector of T^n belonging to eigenvalue c^n.

7. If an n × n matrix A is similar to an n × n matrix B, show that A and B have the same eigen values.

8. Let $f: C^2 \to C^2$ be a linear operator defined by f(x, y) = (3x - y, 13 × -3y). Find all the eigen values and linearly independent eigen vectors. Note that C represents set of complex numbers.

9. If $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, compute A^n where n is an integer.

10. If V is a vector space of functions with basis {cos H, sin θ}.

(i) Find the characteristic polynomial of D.

(ii) Show that D is a zero of the function $f(x) = x^2 + 1$.

11. If A is an invertible matrix over a field K, if λ is an eigen value of A, show that λ^{-1} is an eigen value of A^{-1}.

12. If A is an r × r matrix and P is an invertible matrix then show by induction on n that $(P^{-1}AP)^n = P^{-1}A^nP$.

13. If f(x) is any polynomial over a field, show that $f(P^{-1}AP) = P^{-1}f(A)P$.

14. If f(x) is a polynomial over a field, show that $f(A^t) = f(A))^t$.

15. If A is a symmetric matrix then show that f(A) is also a symmetric matrix for any polynomial f(x).

CHAPTER SEVEN

LINEAR FUNCTIONAL AND DUAL SPACE

7.1 Definition and Examples

Let V be a vector space over a field K.

Notice that K itself is a vector space over itself generated by the multiplicative identity 1. This means that the field K is a vector space over itself and is of dimension 1.

Definition 7.1.1

Let V be a finite dimensional vector space over a field K. A linear map $f : V \to K$ from the vector space into its base field K is called a linear functional.

Example 7.1.1

Let K be a field then K^n is a vector space over K of dimension n. Define $\varphi : K^n \to K$ by

$$\varphi(a) = a_1 + a_2 + \ldots + a_n = \sum_{i=1}^{n} a_i$$

where $a = (a_1, a_2, \ldots, a_n)$ Show φ is a linear functional on K^n.

Solution

If $a \in K^n, b \in K^n$ and c, d $\in$ K let $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$

$$\begin{aligned} \varphi(ca+db) &= \varphi\ (ca_1 + db_1, ca_2 + db_2, \ldots, ca_n + db_n) \\ &= (ca_1 + db_1) + (ca_2 + db_2) + \ldots + (ca_n + db_n) \\ &= (ca_1 + ca_2 \ldots + ca_n) + (db_1 + db_2 + \ldots + db_n) \\ &= c(a_1 + a_2 + \ldots + a_n) + d(b_1 + b_2 + \ldots + b_n) \\ &= c\sum_{i=1}^{n} a_i + d\sum_{i=1}^{n} b_i = c\varphi(a) + d\varphi(b) \end{aligned}$$

hence φ is a linear functional.

Example 7.1.2

Let V be set of all n × n matrices over a field K, we define a map $T : V \to K$ by

$$T(A) = \text{trace}\ (A) = \sum_{i=1}^{n} a_{ii}$$

where $A = (a_{ij})$. Show that T is a linear functional.

Solution

Let A, B ε V where A = (a_{ij}) and B = (b_{ij}) i=1,2,...,n and c, d ε K.

$$cA = (ca_{ij})$$

$$dB = (db_{ij})$$

$$T(cA + dB) = \text{trace } (cA + dB) = \sum_{i=1}^{n} (ac_{ii} + db_{ii})$$

$$= \sum_{i=1}^{n} ca_{ii} + \sum_{i=1}^{n} db_{ii} = c\sum_{i=1}^{n} a_{ii} + d\sum_{i=1}^{n} b_{ii} = c\,T(A) + d\,T(B)$$

T is a linear map hence a linear functional. Let V be a vector space over a field K. Consider the set of all linear functional on V. If T and F are linear functionals, we define addition and scalar multiplication on this set by

$$(T + F)(u) = T(u) + F(u)$$

and

$$(cT)(u) = c(T(u)).$$

The set of all linear functional on V is denoted by V^*.

Theorem 7.1.1

The set V^* of all linear functional on V is a vector space over K.

Proof

Let T, F ε V^*, for u,v ε V and a, b ε K then by definition,

$$(T + F)(au + bv) = T(au + bv) + F(au + bv) = a\,T(u) + bT(v) + aF(u) + bF(v)$$
$$= aT(u) + aF(u) + bT(v) + vF(v) = a(T(u) + F(u)) + b(T(v) + F(v))$$
$$= a(T + F)(u) + b(T + F)(v)$$

T + F is therefore a linear functional on V, hence T + F ε V^*

Similarly for c ε K,

$$(cT)(au + bv) = c(T(au + bv)) = c(aT(u) + bT(v))$$
$$= caT(u) + cbT(v) = a(cT) + b(cT)(v)$$

hence c T ε V^*.. If T, F, H ε V^* then

$$\big((T+F)+H\big)(u) = \big(T+F\big)(u) + H(u) = T(u) + F(u) + H(u)$$

and

$$\big((T+(F+H)\big)(u) = T(u) + (F(u) + H(u)$$

This shows $(T+F)+H=T+(F+H)$ hence associativity of addition holds in V^*. Zero map is the identity while if T ε V^* then -T is the inverse of T.

Also c(T + F) = cT + cF and if r ε K then

$$(c + r)T = cT + rT$$

and so the distributive laws are also satisfied in V^*. This shows that V^* is a vector space over K. V^* is called the Dual space of V.

7.2 Dual Basis

Let V and W be vector spaces over a field K. If dim V = n and dim W = m, recall that Hom (V, W), which is the set of all linear maps from V into W, is also a vector space over K and its dimension is nm. As discussed earlier, K is also a vector space over itself of dimension 1. It follows therefore that the dimension of the dual space $V^* = \text{Hom}(V, K)$ is n.1 = n.

This shows that V^* and V have the same dimension. We will now show that a basis of V determines a basis of V^*.

Theorem 7.2.1

Let V be an n-dimensional vector space over a field K and $(v_1, v_2, \ldots, v_n)$ a basis of V. Let $\phi_1, \phi_2, \ldots, \phi_n \,\varepsilon\, V^*$ defined by

$$\phi_i(v_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \text{ ne } j \end{cases}$$

then $\{\phi_1, \phi_2, \ldots, \phi_n\}$ is a basis for V^*.

Proof:

We need to show that $\{\phi_1, \phi_2, \ldots, \phi_n\}$ generate V^* and are linearly independent. To show that these set of linear functional generate V^*, let T ε V^*. Since $\{v_1, v_2, \ldots, v_n\}$ form a basis for V, let $T(v_1) = c_1, T(v_2) = c_2, \ldots, T(v_n) = c_n$. We want to prove that

$$T = c_1\phi_1 + c_2\phi_2 \ldots + c_n\phi_n.$$

If two functions are equal on the basis of V then they are equal on every element of V. We now show that T and $c_1\phi_1 + c_2\phi_2 + \ldots + c_n\phi_n$ are equal on each of the basis $v_1\, v_2, \ldots, v_n$.

$$\begin{aligned}(c_1\phi_1 + c_2\phi_2 + \ldots + c_n\phi_n)(v_1) &= c_1\phi_1(v_1) + c_2\phi_2(v_1) + \ldots + c_n\phi_n(v_1) \\ &= c_1.1 + c_2.0 + \ldots + c_n.0 = c_1 = T(v_1)\end{aligned}$$

Also

$$\begin{aligned}(c_1\phi_1 + c_2\phi_2 + \ldots + c_n\phi_n)(v_2) &= c_1\phi_1(v_2) + c_2\phi_2(v_2) + \ldots + c_n\phi_n(v_2) \\ &= c_1.0 + c_2.1 + \ldots + c_n.0 = c_2. = T(v_2)\end{aligned}$$

Similarly for i = 3, 4, 5,..., n

$$\begin{aligned}(c_1\phi_1 + c_2\phi_2 + \ldots + c_i\phi_i + \ldots + c_n\phi_n)(v_i) &= c_1\phi_1 + c_2\phi_2(v_i) + \ldots + c_n\phi_n(v_i) \\ &= c_1.0 + c_2.0 + \ldots + c_i.1 + \ldots + c_n.0 \\ &= c_i = T(v_i)\end{aligned}$$

This shows that T and $c_1\phi_1 + c_2\phi_2 + \ldots + c_n\phi_n$ are equal on the basis hence in general we have

$$T = c_1\phi_1 + c_2\phi_2 + \ldots + c_n\phi_n.$$

Therefore, $\{\phi_1 \phi_2, \ldots, \phi_n\}$ generate V^*.

We now show that $\{\phi_1, \phi_2, \ldots, \phi_n\}$ are linearly independent.

Suppose $a_1\phi_1 + a_2\phi_2 + \ldots + a_n\phi_n = 0$ we then apply both sides of the above equation to $(v_1, v_2, \ldots v_n)$ one by one.

$$0 = 0(v_1) = (a_1\phi_1 + a_2\phi_2 + \ldots + a_n\phi_n)(v_1)$$
$$= a_1\phi_1(v_1) + a_2\phi_2(v_1) + \ldots + a_n\phi_n(v_1) = a_1$$

also

$$0 = 0(v_2) = (a_1\phi_1 + a_2\phi_2 + \ldots + a_n\phi_n)(v_2)$$
$$= a_1\phi_1(v_2) + a_2\phi_2(v_2) + \ldots + a_n\phi_n(v_2) = a_1.0 + a_2.1 + \ldots + a_n.0 = a_2$$

Similarly, for $i = 3, 4, 5, \ldots, n$

$$0 = 0(v_i) = (a_1\phi_1 + a_2\phi_2 + \ldots + a_n\phi_n)(v_i)$$
$$= a_1\phi_1(v_i) + a_2\phi_2(v_i) + \ldots + a_i\phi_i(v_i) + \ldots + a_n\phi_n(v_i)$$
$$= a_1.0 + a_2.0 + \ldots + a_i.1 + \ldots + a_n.0 = a_i$$

This proves that $a_1 = 0, a_2 = 0, \ldots, a_n = 0$ hence . $\{\phi_1, \phi_2, \ldots, \phi_n\}$ are linearly independent. Since they also generate V^* then they form a basis for V^*.

Example 7.2.1

Consider the following basis of R^3 $\{v_1 = (1,1,1), v_2 = (1,2,3), v_3 = (2,-1,1)\}$, an find the dual basis $\{\phi_1, \phi_2, \phi_3\}$.

Solution

We want the functionals

$\phi_1(x, y, z) = a_1x + a_2y + a_3z$, $\phi_2(x, y, z) = b_1x + b_2y + b_3z$ and

$\phi_3(x, y, z) = c_1x + c_2y + c_3z$. such that

$\phi_1(v_1) = 1, \phi_1(v_2) = 0, \phi_1(v_3) = 0$

$\phi_2(v_1) = 0\ \phi_2(v_2) = 1, \phi_2(v_3) = 0$

$\phi_3(v_1) = 0\ \phi_3(v_2) = 0, \phi_3(v_3) = 1$

we solve for ϕ_1 by first using the above

$\phi_1(v_1) = 1, \phi_1(v_2) = 0$ and $\phi_1(v_3) = 0$.

$\phi_1(v_1) = \phi_1(1,1,1) = a_1 + a_2 + a_3 = 1$

$\phi_1(v_2) = \phi_1(1,2,3) = a_1 + 2a_2 + 3a_3 = 0$

$\phi_1(v_3) = \phi_1(2,-1,1) = 2a_1 - a_2 + a_3 = 0$

On solving the above simultaneous equation, we have

$$a_1 = 1, a_2 = 1 \text{ and } a_3 = -1$$

Hence

$$\phi_1(x, y, z) = x + y - z$$

We next find ϕ_2 by using the fact that

$$\phi_2(v_1)=0, \phi_2(v_2)=1 \text{ and } \phi_2(v_3)=0$$

$$\phi_2(v_1)=\phi_2(1,1,1)=b_1+b_2+b_3=0$$

$$\phi_2(v_2)=\phi_2(1,2,3)=b_1+2b_2+3b_3=1$$

$$\phi_2(v_3)=\phi_2(2,-1,1)=2b_1-b_2+b_3=0$$

On solving the above equations simultaneously, we have

$$b_1=\frac{-2}{5}, b_2=\frac{-1}{5}, b_3=\frac{3}{5}$$

hence

$$\phi_2(x,y,z)=\frac{-2}{5}x\frac{-1}{5}y+\frac{3}{5}x$$

Finally to find ϕ_3 we have

$$\phi_3(v_1)=\phi_3(1,1,1)=c_1+c_2+c_3=0$$

$$\phi_3(v_2)=\phi_3(1,2,3)=c_1+2c_2+3c_3=0$$

$$\phi_3(v_3)=\phi_3(2,-1,1)=2c_1-c_2+c_3=1$$

On solving the above equations we have

$$c_1=\frac{1}{5}, c_2=\frac{-2}{5} \text{ and } c_3=\frac{1}{5}$$

hence

$$\phi_3(x,y,z)=\frac{1}{5}x\frac{-2}{5}y+\frac{1}{5}z.$$

The next theorem establishes a relationship between the basis of a finite dimensional vector space V and the basis of its dual space V^*.

Theorem 7.2.2

Let V be an n-dimensional vector space with a basis $\{v_1, v_2,..., v_n\}$. Let V^* be the dual space with basis $\{\phi_1, \phi_2,..., \phi_n\}$. If $T \varepsilon V^*$ then $T=T(v_1)\phi_1+T(v_2)\phi_2+...+T(v_n)\phi_n$ and if $u \varepsilon V$ then $u=\phi_1(u)y_1+\phi_2(u)v_2+...+\phi_n(u)v_n$

Proof

Since $T \varepsilon V^*$ and $\{\phi_1, \phi_2,..., \phi_n\}$ is a basis for V^* then $T=a_1\phi_1+a_2\phi_2+...+a_n\phi_n$ for some scalars $a_1, a_2,..., a_n \varepsilon K$. We need to find the values of $a_1 a_2,..., a_n$. It is enough to find the values of T on the basis of V.

$$T(v_1) = (a_1\phi_1+a_2\phi_2+...+a_n\phi_n(v_1) = a_1\phi_1(v_1)+a_2\phi_2(v_1)+...+a_n\phi_n(v_1)$$

$$= a_1.1+a_2.0+...+a_n.0 = a_1$$

hence

$$T(v_1) = a_1$$

$$T(v_2) = (a_1\phi_1 + a_2\phi_2 + \ldots + a_n\phi_n)(v_2) = a_1\phi_1(v_2) + a_2\phi_2(v_2) + \ldots + a_n\phi_n(v_2)$$

$$= a_1.0 + a_2.1 \ldots + a_n.0 = a_2$$

hence

$$T(v_2) = a_2$$

For any (i)

$$T(v_i) = (a_1\phi_1 + a_2\phi_2 + \ldots + a_i\phi_i + \ldots + a_n\phi_n)(v_i)$$

$$= a_1\phi_1(v_i) + a_2\phi_2(v_i) + \ldots + a_i\phi_i(v_i) + \ldots + a_n\phi_n(v_i)$$

$$= a_1.0 + a_2.0 + \ldots + a_i.1 + \ldots + a_n.0$$

$$= 0 + 0 \ldots + a_i + \ldots + 0 = a_i$$

This shows that

$$a_1 = T(v_1).$$

$$a_2 = T(v_2)$$

$$\vdots$$

$$a_n = T(v_n)$$

hence

$$T = a_1\phi_1 + a_2\phi_2 + \ldots + a_n\phi_n \quad T(v_1)\phi_1 + T(v_2)\phi_2 + \ldots + T(v_n)\phi_n.$$

Also since $u \varepsilon V$ and $\{v_1, v_2, \ldots, v_n\}$ is a basis for V then there exists uniquely scalars $c_1, c_2, \ldots, c_n \varepsilon K$ such that $u = c_1 v_1 + c_2 v_2 + \ldots + c_n v_n$.

To determine the image of a vector u under any $T \varepsilon V^*$, , it is enough to determine its image under the basis $\phi_1, \phi_2, \ldots, \phi_n$ hence

$$\phi_1(u) = \phi_1(c_1 v_1 + c_2 v_2 + \ldots + c_n v_n) \quad \phi_1(c_1 v_1) + \phi_1(c_2 v_2) + \ldots + \phi_1(c_n v_n).$$

$$= c_1\phi_1(v_1) + c_2\phi_1(v_2) + \ldots + c_n\phi_1(v_n) = c_1.1 + c_2.0 + \ldots + c_n.0 \quad c_1$$

Similarly

$$\phi_2(u) = \phi_2(c_1 v_1 + c_2 v_2 + \ldots + c_n v_n) \quad \phi_2(c_1 v_1) + \phi_2(c_2 v_2) + \ldots + \phi_2(c_n v_n)$$

$$= c_1\phi_2(v_1) + c_2\phi_2) + \ldots + c_n\phi_2(v_n) = c_1.0 + c_2.1 + \ldots + c_n.0 = c_2$$

$$\phi_i(u) = \phi_i(c_1 v_1 + c_2 v_2 + \ldots + c_i v_i + \ldots + c_n v_n)$$

$$= \phi_i(c_1 v_1) + \phi_i(c_2 v_2) + \ldots + \phi_i(c_i v_i) + \ldots + \phi_i(c_n v_n)$$

$$= c_i\phi_i(v_1) + c_2\phi_i(v_2) + \ldots + c_i\phi_i(v_i) + \ldots + c_n\phi_i(v_n)$$

$$= c_1.0 + c_2.0 + \ldots + c_i.1 + \ldots + c_n.0 = c_i$$

hence

$$c_1 = \phi_1(u)$$
$$c_2 = \phi_2(u)$$
$$\vdots$$
$$c_n = \phi_n(u)$$

It then follows that

$$u = c_1 v_1 + c_2 v_2 + \ldots + c_n v_n = \phi_1(u)v_1 + \phi_2(u)v_2 + \ldots + \phi_n(u)v_n$$

Let V be a finite dimensional vector space, and V^* its dual space. Since V^* is a vector space, it also has a dual space called the second dual space of V. This second dual of V is the set of all linear functional on V^* and it is denoted by V^{**}.

Theorem 7.2.3

Each vector v ε V determines a specific element $\hat{v}$ ε V^{**}.

Proof

We need to show that if we have a vector v ε V then there exists a linear functional $\hat{v}$ ε V^{**}. We define $\hat{v}(\varphi) = \varphi(v)$ for φ ε V^* and v ε V. We need to show that this map $\hat{v}$: $V^* \to K$ is linear. Let $\phi_1, \phi_2 \, \varepsilon \, V^*$ and a, b ε K

$$\hat{v}\,(a\phi_1 + b\phi_2) = (a\phi_1 + b\phi_2)(v) = a\phi_1 + b\phi_2(v) = a\hat{v}(\phi_1) + b\hat{v}(\phi_2)$$

hence $\hat{v}$ is linear and so $\hat{v}$ ε V^{**}. Since V is finite dimensional then if dim (V) = n, we have

$$n = \dim (V) = \dim \;(V^*)$$

Also $V^{**} = \mathrm{Hom}(V^*, K)$ and since dim (K) = 1 then

$$\dim(V^{**}) = \dim V^* \dim K = \dim(V^*).$$

It then follows that

$$\dim (V) = \dim \;(V^*) = \dim(V^{**}).$$

By theorem 7.2.3 a specific vector v ε V determines a specific linear functional $\hat{v}$ ε V^{**} and so the natural mapping $v \to \hat{v}$ determines an isomorphism between V and V^{**}. Accordingly, we can view V as a space of linear functional on V^* and write $V = V^{**}$.

If $\{\phi_1, \phi_2, \ldots, \phi_n\}$ is the basis for V^* and $\{v_1, v_2, \ldots, v_n\}$ a basis of V then $\{v_1, v_2, \ldots, v_n\}$ is a basis of $V = V^{**}$ which is dual to $\{\phi_1, \phi_2, \ldots, \phi_n\}$.

7.3 Annihilators and Transpose of a Linear Map

Definition 7.3.1

Let V be a finite dimensional vector space over a field K and W a subset of V. A linear functional $\varphi \,\varepsilon\, V^*$ is called an annihilator of W if $\varphi(w)=0$ for all $w \,\varepsilon\, W$. This means that $\varphi \,\varepsilon\, V^*$ is an annihilator of W if $\varphi(W) = 0$. The set of all linear functionals which annihilate W is denoted by W° and it is called annihilator of W.

Theorem 7.3.1

Let V be a finite dimensional vector space over a field K and W a subset of V. The annihilator W° of W is a subspace of V^*.

Proof:

Let O be the zero linear functional on V then $O(w) = 0$ for all $w \,\varepsilon\, W$. This shows that $O \,\varepsilon\, W^\circ$ hence W° is not empty.

If $\phi_1, \phi_2 \,\varepsilon\, W^\circ$ and $a, b \,\varepsilon\, K$, then $\phi_1(w)=0$ and $\phi_2(w)=0$ for all $w \,\varepsilon\, W$. We then have

$$(a\phi_1+b\phi_2)(w)=a\,\phi_1(w)+b\,\phi_2(w)=a.0+b.0=0$$

hence $a\phi_1+b\phi_2 \,\varepsilon\, W^\circ$. This shows that W° is a subspace of V^*.

In the definition for annihilator of W, W need not be a subspace but if W is a subspace of V, then we have the following theorem.

Theorem 7.3.2

If V is a finite dimensional vector space and W is a subspace of V then

(i) $\dim V = \dim W + \dim W^\circ$.

(ii) $W = W^{00}$

where $W^{00}=\left(W^\circ\right)^0=\{v \,\varepsilon\, W : \phi(v)=o \text{ for every } \varphi \,\varepsilon\, W^\circ\}$. W^{00} is viewed as a subspace of V under the identification of V and V^{**}.

Proof:

(i) Let $\dim V = n$ and $\dim W = r \le n$. To show that

$$\dim V = \dim W + \dim W^\circ,$$

we need to show that $\dim W^\circ = n - r$. Let $\{w_1, w_2, \ldots, w_r\}$ be a basis for W. Since W is a subspace of V, its basis can be extended to a basis of V, namely $\{w_1, \ldots, w_r, v_1\, v_2, \ldots v_{n-r}\}$.

By Theorem 7.2.1, to every basis of V corresponds a dual basis of V^* Let $\{\phi_1, \phi_2, \ldots, \phi_r, \psi_1, \ldots, \psi_{n-r}\}$ be the dual basis corresponding to the basis $\{w_1, \ldots, w_r, v_1, v_2 \ldots v_{n-r}\}$.

By the definition of dual basis, $\psi_i(w_j)=0$ for $i = 1,2,3,\ldots,n-r$ and $j=1, 2, \ldots, r$.

This shows that each ψ_i annihilates each w_j and since the w_js form the basis for W then each ψ_i annihilates every element of W.

This shows that $\psi_1, \psi_2, \ldots, \psi_{n-r} \, \varepsilon \, W^\circ$. We now show that $\{\psi_1, \psi_2, \ldots, \psi_{n-r}\}$ is actually a basis since $\{\phi_1, \phi_2, \ldots, \phi_r, \psi_1, \ldots \psi_{n-r}\}$ is a basis for V^*, then $\{\psi_1, \psi_2, \ldots, \psi_{n-r}\}$ is linearly independent being part of a basis, we therefore only need to show that they generate W°.

If $\psi \; \varepsilon \; W^\circ$, then by theorem 7.2.2

$$\psi = \psi(w_1)\phi_1 + \ldots + \psi(w_r)\phi_r + \psi(v_1)\psi_1 + \ldots + \psi(v_{n-r})\psi_{n-r}$$

Since $\psi \; \varepsilon \; W^\circ$ then $\psi(w_i) = 0$ for i = 1, 2,...,r then

$$\psi = 0.\phi_1 + \ldots + 0.\phi_r + \psi(v_1)\psi_1 + \ldots + \psi(v_{n-r})\psi_{n-r} = \psi(v_1)\psi_1 + \ldots + \psi(v_{n-r})\psi_{n-r}$$

This shows that $\{\psi_1, \psi_2, \ldots, \psi_{n-r}\}$ generate W° and so it is a basis of W°. We have

$$\dim W^\circ = n - r = \dim V - \dim W$$

hence

$$\dim W^\circ + \dim W = \dim V.$$

(ii) dim V = n and dim W = r, since dim V = dim V^* then

$$\dim V^* = n$$

and by (i)

$$\dim W^\circ = n - r.$$

Applying (i) to the fact that $W^{infinity} = (W^\circ)^\circ$, then we have

$$\dim W^{00} = \dim V - \dim W^0 = n - (n - r) = r = \dim W$$

hence

$$W = W^{00}.$$

Definition 7.3.2

Let U and V be vector spaces over a field K and T: $U \rightarrow V$ a linear map from U into V. If $\varphi \; \varepsilon \; V^*$ then $\phi : V \rightarrow K$ is a linear functional on V then the composite map $\varphi.T \; \varepsilon \; V^*$ is a linear map

$$\begin{tikzcd} U \arrow[r, "T"] & V \arrow[r, "\phi"] & K \end{tikzcd}$$

We therefore see that the correspondence $\varphi \rightarrow \varphi.T$ is a mapping from V^* into U^* and it is called the transpose of T and it is denoted by T^t. If given a linear map T: U → V then this induces a map $T^t : V^* \rightarrow U^*$ where $T^t(\phi) = \phi.T$ where $\varphi \; \varepsilon \; V^*$.

Theorem 7.3.3

Given a linear map $T : U \rightarrow V$, the induced map $T^t : V^* \rightarrow U^*$ defined by $T^t(\phi) = \phi.T$ is a linear map.

Proof:

Let $\phi_1, \phi_2 \varepsilon V^*$ and a, b, ε K then

$$T'(a\phi_1+b\phi_2) = (a\phi_1+b\phi_2).T = a(\phi_1.T)+b(\phi_2.T) = aT'(\phi_1)+bT'(\phi_2)$$

T' is a linear map.

Example 7.3.1

Let V be a finite dimensional vector space over a field K and W a subset of V. If L(W) represents the set of all vectors in V generated by W. Show that $W^\circ = (L(W))^\circ$.

Solution

We show that $\varphi \varepsilon W^\circ$ if and only if $\varphi \varepsilon (L(W))^\circ$. If $\varphi \varepsilon W^\circ$ then $\varphi \varepsilon V^*$ and $\varphi(w) = 0$ for all $w \varepsilon W$. If $v \varepsilon L(W)$ then there exists $w_1, w_2, \ldots, w_r \varepsilon W$ and $a_1, a_2, \ldots, a_r \varepsilon K$ such that $v = a_1w_1 + a_2w_2 + \ldots + a_rw_r$, then

$$\phi(v) = \phi(a_1w_1+a_2w_2+\ldots+a_rw_r) = a_1\phi(w_1)+a_2\phi(w_2)+\ldots+a_r\phi(w_r)$$
$$= a_1.0+a_2.0+\ldots+a_r.0 = 0+0+\ldots+0=0.$$

This shows that $\phi \varepsilon (L(W))^\circ$.

Example 7.3.2

If $\phi: R^2 \to R$ is a linear functional defined by $\varphi(x, y) = 2x - 3y$. If $T: R^3 \to R^2$ is a linear map defined by $T(x, y, z) = (x + y, y + z)$. Find $(T'(\phi))(x,y,z)$.

Solution

By definition, $T'(\phi) = \phi.T$, hence

$$(T'(\phi))(x,y,z) = \phi(T(x,y,z)) = \varphi(x+y; y+x)$$
$$= 2(x+y) - 3(y+z) = 2x+2y-3y-3z = 2x-y-3z$$

7.4 Exercise

1. Let $u_1 = (1,1,1)\, u_2 = (1,1,0), u_3 = (1,0,0)$ be basis for R^3, compute the dual basis.
2. If U and W are subset of a vector space V over a field K, show that if $U \subset W$ then $W^\circ \subset U^\circ$.
3. If U and W are subspaces of a vector space V then show that $(U+W)^\circ = U^\circ \cap W^\circ$
4. If V, U, W are vector spaces over a field K and F: V → U, T: U → W are linear maps. Show that $(T.F)' = F'.T'$.
5. If W is a subspace of R^4 spanned by the vectors $w_1 = (1,4,-1,8)$, $w_2 = (1,2,-3,4)$, $w_3 = (1,3,-2,4)$. Find the basis of the annihilator of W.
6. If $u_1 = (2, -4, 7)$, $u_2 = (1, -1, 1)$ and $= (1, -2, 3)$ are vectors in $\Re^3$

(i) show that u_1, u_2, u_3 from a basis for R^3

(ii) find the dual basis for $\{u_1, u_2, u_3\}$

7. If V is a vector space, U and W subspaces of V.

(i) Let $W^\circ = \{\varphi(w)\ V^* \mid \varphi(w) = 0$ for all $w \varepsilon W\}$. Show that W° is a subspace of V^*

(ii) Show that $(U \cap W)^\circ = U^\circ + W^\circ$.

8. W is a subspace of R^4 generated by vectors $u_1 = (1, -2, 5, -3)$ u_2 $(2, 3, 1, -4)$ and $u_3 = (3, 8, -3, 5)$.

(i) Find the basis and dimension of W

(ii) Find the basis for the annihilator W° of W.

9. If φ is a linear functional on R^2 defined by $\varphi(2, 1) = 15$ and $\varphi(-2, 7) = 41$ then

(i) find $\varphi(x, y)$ (ii) find $\varphi(1, -2)$.

10. Let V be a vector space over R. If u, v, ε V, the line segment joining the vectors u, v, which is denoted by $\overline{uv}$ is defined as $\overline{uv} = \{t(u) + 1-t)v$ where $0 \leq t \leq 1\}$.

Also a subset S of V is said to be convex if whenever u, v ε S then $\overline{uv} \subset S$. If φt V^* and $W^+ = \{v \varepsilon V \mid \varphi(v) > 0\}$ and $W = \{\varphi\ v \varepsilon V \mid \varphi(v) = 0\}$, show that W- and W a convex.

11. Let φ and γ be linear functionals on R^3 defined thus, $\varphi(x, y, z) = 4x - 2y + 3z$ and $\gamma(x, y, z) = 8x - 12y + 4z$. Find (i) $3\varphi + 2\gamma$ (ii) $5\varphi - 2\gamma$.

12. Let V be a vector space of polynomials over a field F. Let b ε F, we define a mapp $\varnothing_b : \forall \text{—>} f$ by $\phi_b(f(x)) = f(a)$

(i) show that $\varnothing_b$ is a linear functional

(ii) show if $b \underline{=} c$ then $\phi_b \neq \phi_c$.

CHAPTER EIGHT

BILINEAR FORMS

8.1 Definition and Examples

Definition 8.1.1

Let V be a vector space over a field K. A bilinear form on V is a function f which assigns to each ordered pair of vectors u, v in V a scalar f(u,v) ε K and which satisfies the following axioms.

(i) $f(au + bw, v) = af(u,v) + bf(w,v)$

(ii) $f(u, cw + dv) = cf(u,w) + df(u,v)$

where u, w, v ε V and a,b,c,d ε K.

If we recall that V x V denotes the set of ordered pairs of vectors in V then the above definition can be rephrased as follows:

A bilinear form on V is a function, $f: V \times V \rightarrow K$ from $V \times V$ into K such that it is linear in the first and second coordinates.

Example 8.1.1

Let $f: R^2 x R^2 \rightarrow R$ be defined by $f(u, v) = 2u_1v_2 - 3u_2v_1$ where $u = (u_1, u_2)$, $v = (v_1, v_2)$ show that f is bilinear.

Solution

Let $u = (u_1, u_2)$, $w = (w_1, w_2)$ and $v = (v_1, v_2)$ be vectors in R^2 and $a, b, c, d, \varepsilon R$.

(i)
$$
\begin{aligned}
f(au+bw, v) &= f((au_1 + bw_1, au_2 + bw_2), (v_1, v_2)) \\
&= 2(au_1 + bw_1) v_2 - 3(au_2 + bw_2) v_1 \\
&= 2au_1v_2 + 2bw_1v_2 - 3au_2v_1 - 3bw_2v_1 \\
&= 2au_1v_2 - 3au_2v_1 + 2bw_1v_2 - 3bw_2v_1 \\
&= a(2u_1v_2 - 3u_2v_1) + b(2w_1v_2 - 3w_2v_1) = af(u, v) + bf(w, v)
\end{aligned}
$$

hence f is linear in the first coordinate.

(ii)
$$
\begin{aligned}
f(u, cw+dw) &= f((u_1, u_2), (cw_1 + dv_1, cw_2 + dv_2)) \\
&= 2u_1(cw_2 + dv_2) - 3(u_2(cw_1 + dv_1)) \\
&= 2u_1 cw_2 + 2u_1 dv_2 - 3u_2 cw_1 - 3u_2 dv_1 \\
&= 2u_1 cw_2 - 3u_2 cw_1 + 2u_1 dv_2 - 3u_2 dv_1 \\
&= c(2u_1 w_2 - 3u_2 w_1) + d(2u_1 v_2 - 3u_2 v_1) = cf(u, w) + df(u, v)
\end{aligned}
$$

f is therefore linear in the second coordinate hence f is bilinear.

Example 8.1.2

Let $\phi_1 : V \to K$ and $\phi_2 : V \to K$ be linear functionals on V. We define a map $f: V \times V \to K$ by $f(u,v) = \phi_1(u)\phi_2(v)$. Show f is bilinear.

Solution

If u, w, v ε V and a, b, c, ε K since ϕ_1 and ϕ_2 are linear functionals, then

$$\phi_1(au+bw) = a\phi_1(u)+b\phi_1(w)$$

and

$$\phi_2(cw+dv) = c\phi_2(w)+d\phi_2(v).$$

It then follows that

$$\begin{aligned} f(au+bw, v) = \phi_1(au+bw)\phi_2(v) &= (a\phi_1(u)+b\phi_1(w))\phi_2(v) \\ &= a\phi_1(u)\phi_2(v)+b\phi_1(w)\phi_2(v) = af(u,v)+bf(w,v) \end{aligned}$$

This shows f is linear in the first coordinate. Also

$$\begin{aligned} f(u, cw+dv) = \phi_1(u)\phi_2(cw+dv) &= \phi_1(u)(c\phi_2)w)+d\phi_2(v)) \\ &= \phi_1(u)c\phi_2(w)+\phi_1(u)d\phi_2(v) = c\phi_i(u)\phi_2(w)+d\phi_1(u)\phi_2(v) \\ &= cf(u,w)+df(u,v) \end{aligned}$$

f is linear in the second coordinate hence it is bilinear. Let V be an n-dimensional vector space over a field K. If $f: V \times V \to K$ is a bilinear map. Let u, v, ε V, if $(e_1, e_2, \ldots e_n)$ is a basis for V, then

$$u = a_1e_1+a_2e_2+\ldots+a_ne_n$$

and

$$v = b_1e_1+b_2e_2+\ldots+b_ne_n$$

where a_is and b_is are scalars in K.

$$\begin{aligned} f(u,v) &= f(a_1e_1+a_2e_2+\ldots+a_ne_n, b_1e_1+b_2e_2+\ldots+b_ne_n) \\ &= a_1f(e_1, b_1e_1+b_2e_2+\ldots+b_ne_n)+a_2f(e_2, b_1e_1+\ldots+b_ne_n)+\ldots \\ &\quad + a_nf(e_n, b_1e_1+\ldots+b_ne_n) \\ &= a_1b_1f(e_1,e_1)+a_1b_2f(e_1,e_2)+\ldots+a_1b_nf(e_1,e_n) \\ &\quad + a_2b_1f(e_2,e_1)+\ldots+a_2b_nf(e_2,e_n)+\ldots \\ &\quad + a_nb_1f(e_n,e_n)+\ldots+a_nb_nf(e_n,e_n) \\ &= \sum_{i,j=1}^{n} a_ib_jf(e_i,e_j). \end{aligned}$$

The bilinear form f is completely determined by the values $f(e_i, e_j)$.

If $a_{ij} = f(e_i, e_j)$ then the matrix $A = (a_{ij})$ is called the matrix representing f relative to the basis $\{e_i\}$.

The matrix $A = (a_{ij}) = (f(e_i, e_j))$ is sometimes simply referred to as the matrix of f in the basis $\{e_i\}$.

The matrix $A = (a_{ij}) = (f(e_i, e_j))$ represents the bilinear form f in the following sense, since

$$u = a_1 e_1 + a_2 e_2 + ... + a_n e_n$$

and

$$v = b_1 e_1 + ... + b_n e_n .$$

then

$$[u]_e = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix} \text{ and } [v]_e = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix},$$

where $e = (e_1, e_2, ..., e_n)$

$$f(u, v) = \sum_{i,j=1}^{n} a_i b_j f(e_i, e_j) = (a_1, ..., a_n) A \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{pmatrix} = [u]_e^t A [v]_e .$$

Example 8.1.3

Consider the map $f: R^2 x R^2 \rightarrow R$ defined by $f(u, v) = 2u_1 v_2 - 3u_2 v_1$ where $u = (u_1, u_2), v = (v_1 v_2)$, compute the matrix of f in the basis $e_1 = (2, 1), e_2 = (1, -1)$.

Solution

Since the bilinear form is on R^2 then the matrix required is a 2×2 matrix.

$a_{11} = f(e_1, e_1) = f((2,1),(2,1)) = 2.2.1 - 3.1.2 = 4 - 6 = 2$

$a_{12} = f(e_1, e_2) = f((2, 1), (1, -1) = 2.2.(-1) - 3.1.1 = -4 - 3 = -7$

$a_{21} = f(e_2, e_1) = f((1, -1), (2, 1)) = 2.1.1 - 3.(-1).2 = 2 + 8 = 8$

$a_{22} = f(e_2, e_2) = f((1, -1), (1, -1)) = 2.1.(-1) - 3.(-1).1 = -2 + 3 = 1$

hence

$$A = \begin{pmatrix} -2 & -7 \\ 8 & 1 \end{pmatrix}$$

We can check if this matrix actually represents the bilinear form A. Let u, v ε R^2 such that $u = a_1 e_1 + a_2 e_2$ and $v = b_1 e_1 + b_2 e_2$ hence

$$u = (2a_1 + a_2, a_1 - a_2) \text{ and } v = (2b_1 + b_2, b_1 - b_2)$$

$$f(u, v) = f((2a_1 + a_2, a_1 - a_2), (2b_1 + b_2, b_1 - b_2))$$

$$= 2(2a_1 + a_2)(b_1 - b_2) - 3(a_1 - a_2)(2b_1 + b_2)$$

$$= 2(2a_1b_1 - 2a_1b_1 + a_2b_1 - a_2b_2) - 3(2a_1b_1 + a_1b_2 - 2a_2b_1 - a_2b_2)$$

$$= 4a_1b_1 - 4a_1b_2 + 2a_2b_1 - 2a_2b_1 - 6a_1b_1 - 3a_1b_2 + 6a_2b_1 + 3a_2b_2$$

$$= -2a_1b_1 + 8a_2b_1 - 7a_1b_2 + a_2b_2$$

However

$$[u]_e = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

and

$$[v]_e = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad = [u]^t_e A \{v\}_e = (a_1, a_2)\begin{pmatrix} -2 & -7 \\ 8 & 1 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

$$= (-2a_1 + 8a_2, -7a_1 + a_2)\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = (2a_1 + 8a_2)b_1 + (-7a_1 + a_2)b_2$$

$$= -2a_1b_1 + 8a_2b_1 - 7a_1b_2 + a_2b_2$$

We have now seen that $f(u, v) = \{u\}^t_e A \{v\}_e$.

We have seen that given a bilinear form f on a vector space V and a basis of the vector space then we can compute a matrix which represents that bilinear form, in that basis. We now want to see how the matrix will transform, when a new basis is selected i.e. if a vector space has two basis e and e' if A is the matrix which represents bilinear form f in the basis e and B is the matrix which represents the Bilinear form f in the basis e' we want to study the relationship between A and B.

Theorem 8.1.1

Let f be a bilinear form on a finite dimensional vector space V. If $e = (e_1, e_2, \ldots, e_n)$ and $e' = (e_{1'}, e_{2'}, \ldots, e_{n'})$ are two basis for V and P is the transition matrix frome $= \{e_i\}$ to $e' = \{e_{i'}\}$. If A is the matrix of f in the basis e and B is the matrix of f in the basis e', then

$$B = p^t A P.$$

Proof :

Since P is transition matrix from basis

$$e = \{e_i\} \text{ to } e' = \{e_i\}$$

then for $u, v \varepsilon V$

$$P[u]_{e'} = [u]_e \text{ and } P[v]_{e'} = [v]_e$$

hence

$$(P[u]_{e'})' = [u]^t_e,$$

this implies $[u]^t_{e'} p' = [u]^t_e$. Thus

$$f(u, v) = [u]^t_e A [v]_e = [u]^t_{e'} p' A P[v]_{e'}.$$

Since u and v are arbitrary elements of V then $P^t A P$ is the matrix of f in the basis e', hence $B = P^t A P$.

Example 8.1.4

Consider $f: R^2 \times R^2 \rightarrow R$ defined by $f(u, v) = u_1 v_1 + u_1 v_2 + u_2 v_1 + u_2 v_2$ where $u = (u_1, u_2), v = (v_1 v_2)$.

Consider the usual basis $e = \{e_1 = (1, 0)\ e_2 = (0, 1)\}$ and another basis $e' = \{e_{1'} = (1, -1), e_{2'} = (1, 1)\}$. In the basis $e = (e_1, e_2)$, matrix $A = (a_{ij})$ where

$$a_{11}, = f((e_1\ e_1) = f(1, 0), (1, 0)) = 1$$
$$a_{12} = f(e_1\ e_2) = f((1, 0), (0, 1)) = 1$$
$$a_{21} = f(e_2\ e_1) = f((0, 1), (1, 0)) = 1$$
$$a_{22} = f(e_2, e_2) = f((0, 1), (0, 1)) = 1$$

hence

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

The matrix B representing f in the basis $e' = (e_{1'}, e_{2'})$ where $B = (b_{ij})$ is

$$b_{11} = f(e_{1'}, e_{1'}) = f((1, 1), (1, -1) = 0$$
$$b_{12} = f(e_{1'}, e_{2'}) = f((1, -1), (1, 1)) = 0$$
$$b_{21} = f(e_{2'}, e_{1'}) = f((1, 1), (1, -1)) = 0$$
$$b_{22} = f(e_{2'}, e_{2'}) = f((1, 1), (1, 1)) = 4$$

hence

$$B = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}$$

The transition matrix P from e' to e is

$$P = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

$$P^t A P = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} = B$$

Definition 8.1.2

An $n \times n$ matrix B is congruent to an $n \times n$ matrix A if there exists a non-singular matrix P such that $B = P^t A P$.

Theorem 8.1.1 therefore implies that all matrices representing the same bilinear form are congruent.

The definition also shows that congruent matrices have the same rank hence if two matrices A and B represent the same bilinear form on V with respect to different basis then they have the same rank. This makes it possible to define the rank of a bilinear form on V as the rank of any matrix which represents the bilinear form with respect to any basis of V. We say that f is degenerate if rank (f) < dim V and non-degenerate if rank (f) = dim V.

8.2 Skew-symmetric and Symmetric Bilinear Forms

Definition 8.2.1

Let f be a bilinear form on a finite dimensional vector space V, f is called an alternating form if $f(u, u) = 0$ for all $u \varepsilon V$.

Theorem 8.2.1

Let f be an alternating bilinear form, then $f(u, v) = -f(v, u)$ for all $u, v \varepsilon V$.

Proof:

Since f is an alternating bilinear form, then

$$f(u + v, u + v) = 0.$$

$$f(u + v, u + v) = f(u, u + v) + f(v, u + v) = f(u, u) + f(u, v) + f(v, u) + f(v, v)$$

but

$$f(u, u) = 0 = f(v, v)$$

then

$$f(u + v, u + v) = f(u, v) + f(v, u).$$

Since $f(u + v, u + v) = 0$, then

$$f(u, v) + f(v, u) = 0.$$

This implies $f(u, v) = -f(v, u)$.

A bilinear form f is called skew-symmetric if $f(u, v) = -f(v, u)$ for all $u, v \varepsilon V$.

Theorem 8.2.2

Let f be a bilinear form on a finite dimensional vector space V then f is skew-symmetric if and only if it is represented by a skew-symmetric matrix A in some ordered basis.

Proof:

Recall that an $n \times n$ matrix A is skew-symmetric if $A^t = -A$. Suppose f is skew-symmetric bilinear form and A represents f in some ordered basis e of V. We need to show that $A^t = -A$. Since f is skew-symmetric then

$$f(u, v) = -f(v, u) \text{ for all } u, v \varepsilon V.$$

Since A is the matrix representing f in some basis e, then

$$f(u, v) = [u]_e^t A [v]_e$$

and

$$f(v,u) = [v]_e^t A [u]_e$$

$f(u, v) = -f(v, u)$ implies $[u]_e^t A [v]_e = -[v]_e^t A [u]_e$.

Since $[u]_e^t A[v]_e$ is a scalar in the field K, then it is equal to its transpose hence

$$[u]_e^t A[v]_e = \left([u]_e^t A[v]_e\right)^t = [v]_e^t A^t [u]_e$$

We have

$$[v]_e^t A^t [u]_e = [u]_e^t A\,[v]_e = -[v]_e^t A[u]_e = [v]_e^t(-A)[u]_e\,.$$

hence

$$A^t = -A$$

Conversely suppose A is skew-symmetric matrix which represent a bilinear form f on V with respect to some basis e of V, we want to show that f(u, v) = -f(v, u) for all u, v ε V. Since A represent f in some basis e, then

$$f(u, v) = [u]_e^t A[v]_e$$

$$f(v, u) = [v]_e^t A[v]_e$$

since $[u]_e^t A[v]_e$ is a scalar, it is equal to its transpose hence

$$f(u, v) = [u]_e^t A[v]_e = \left([u]_e^t A[v]_e\right)^t = [v]_e^t A^t [u]_e$$

by assumption, $A^t = -A$ hence the above equals

$$[v]_e^t(-A)[u]_e = -[v]_e^t A[u]_e = -f(v,u)$$

Hence f(u,v) = -f(v, u) for all u, v ε V and so f is skew-symmetric.
Notice that if f is skew-symmetric, any matrix A representing f in some ordered basis $e = \{e_i\}$ will have its diagonal entries to be zero. This is simply because $f(e_i, e_i) = 0$.

Theorem 8.2.3
Let f be a skew-symmetric bilinear form on an n-dimensional vector space V over a field K. There exists a basis of V in which f can be represented by a matrix of the form.

$$\begin{pmatrix}
0 & 1 & & & & & & & & & \\
-1 & 0 & & & & & & & & & \\
 & & 0 & & & & & & & & \\
 & & & 0 & 1 & & & & & & \\
 & & & -1 & \cdot & & & & & & \\
 & & & & & \ddots & & & & & \\
 & & & & & & 0 & 1 & & & \\
 & & & & & & -1 & 0 & & & \\
 & & & & & & & & 0 & & \\
 & & & & & & & & & \ddots & \\
 & & & & & & & & & & 0
\end{pmatrix}$$

The number of copies of $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is $\frac{1}{2}$ rank (f). This shows that rank (f) is even.

Proof:

Let dim $V = n$, we will prove this theorem by induction on n. If $f = 0$ then the theorem is trivially true because the matrix will be zero matrix. If $n = 1$, then each vector in V will be a scalar multiple of the generator say $u \varepsilon V$. If $a, b \varepsilon K$ then

$$f(au, bu) = abf(u,u) = 0$$

since $f(u,u) = 0$, hence the image of any ordered pair under f is zero therefore $f = 0$ and the theorem is also true. We then assume $n > 1$ and $f = 0$. It then follows that there exists non-zero vectors u_1, u_2 such that $f(u_1, u_2) \neq 0$. We can assume that $f(u_1, u_2) = 1$ because if $f(u_1, u_2) = r$ where $r = 1$ and $r = 0$, then instead of using u_1, we can use $\frac{1}{r}u_1$ to have $f\left(\frac{u_1}{r}, u_2\right) = \frac{1}{r}f(u_1, u_2) = \frac{1}{r}.r = 1$ hence we can assume that $f(u_1, u_2) = 1$. Since f is skew-symmetric, then $f(u_2, u_1) = -1$, u_1, u_2 must be linearly independent because if $u_2 = au_1$ for some $a \varepsilon K$ then

$$f(u_1, u_2) = f(u_1, au_1) = af(u_1, u_1) = 0.$$

Contradiction to the fact that $f(u_1, u_2) \neq 0$.

Let U be the subspace of V generated by u_1 and u_2, then $f: U \times U \rightarrow K$ is represented by the matrix A given by

$$a_{11} = f(u_1, u_1) = 0$$

$$a_{12} = f(u_1, u_2) = 1,$$

$$a_{21} = f(u_2, u_1) = -1$$

and

$$a_{22} = f(u_2, u_2) = 0$$

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

which is the required form for $n = 2$ with basis $\{u_1, u_2\}$.

Let $W = \{w \varepsilon V \mid f(w, u) = 0$ for every $u \varepsilon U\}$. Since $\{u_1, u_2\}$ form a basis for U, if $w \varepsilon$ W then

$$f(w, u) = 0 \text{ for } u \varepsilon U$$

implies $f(w, u_1) = 0$ and $f(w, u_2) = 0$.

We need to show that $V = U \oplus W$. This means that we show that $V = U + W$ and $U \cap W = \{0\}$. We first show that $U \cap W = \{0\}$.

Let $v \varepsilon U \cap W$, then $v \varepsilon U$ and $v \varepsilon W$. Since $v \varepsilon U$ then

$$v = au_1 + bu_2$$

where $a, b \varepsilon K$. Also since $v \varepsilon W$, then

$$f(v, u_1) = 0 \text{ and } f(v, u_2) = 0,$$

then

$$0 = f(v, u_1) = f(au_1 + bu_2, u_1) = af(u_1, u_1) + bf(u_2, u_1)$$
$$= a.0 + bf(u_2, u_1) = bf(u_2, u_1)$$

We therefore have

$$bf(u_2, u_1) = 0 \text{ but } f(u_2, u_1) = -1 \neq 0$$

hence b = 0. Similarly

$$0 = f(v, u_2) = f(au_1 + bu_2, u_2) = af(u_1, u_2) + bf(u_2, u_2)$$
$$= af(u_1, u_2) + b.0 = a(fu_1, u_2)$$

but

$$f(u_1, u_2) = 1$$

hence

$$a = 0.$$

Since a = 0, b = 0 then v = 0, hence $U \cap V = \{o\}$. We next show that V = U + W. Let v ε V and u ε U defined by

$$u = f(v, u_1)u_1 - f(v, u_2)v_2 .$$

We show that w = v - u ε W. It will then follow that v = u + w where u ε U and w ε W. To show that w ε W, we need to show that

$$f(w, u_1) = 0 \text{ and } f(w, u_2) = 0.$$

$$f(u, u_1) = f(f(v, u_1)u_1 - f(v, u_1)u_2, u_1)$$
$$= f(v, u_1)f(u_1, u_1) - f(v, u_1)f(u_2, u_1) = 0 - f(v, u_1)(-1) = f(v, u_1)$$

hence

$$f(v, u_1) = f(u, u_1) .$$

Similarly $f(v, u_2) = f(u, u_2)$ then

$$f(w, u_1) = f(v - u, u_1) = f(v, u_1) - f(u, u_1) = f(u, u_1) - f(u, u_1) = 0.$$

Similarly

$$f(w, u_2) = f(v - u, u_2) = f(v, u_2) - f(u, u_2) = f(u, u_2) - f(u, u_2) = 0.$$

hence w ε W and so

$$V = U + W.$$

We have therefore shown that $V = U \oplus W$. Restriction of f to W is also skew-symmetric form and dim W = n - 2. By induction then the theorem is true for f restricted to W hence there exist basis $(u_3, u_4, \ldots, u_n)$ on which a matrix representing f has the desired form. Thus the basis $(u_1, u_2, u_3 \ldots, u_n)$ of V is the required basis in which the matrix representing f has the desired form. If there are r of these matrices, since the rank of each of these matrices is 2 then the rank of the whole matrix is 2r.

Definition 8.2.2

Let V be an n-dimensional vector space over a field K and f a bilinear form on V. We say that f is symmetric if $f(u, v) = f(v, u)$ for all $u, v \varepsilon V$.

Theorem 8.2.4

If V is finite dimensional vector space over a field K and f is a bilinear form on V then f is symmetric if and only if the matric A representing f in some basis is symmetric.

Proof:

The proof of this theorem follows similar proof as theorem 8.2.2.

Theorem 8.2.5

Let V be an n-dimensional vector space over a field K. If f is a symmetric bilinear form on V, then V has a basis $(v_1, v_2, \ldots, v_n)$ on which f is represented by a diagonal matrix $A = (a_{ij}) = f(v_i, v_j)$ where $f(v_i, v_j) = 0$ for $i = j$.

Proof:

The proof will be similar to the skew-symmetric case and it will be on method of induction on n. If $f = 0$ then the theorem is trivially true because the matrix representing f in any basis will be the zero matrix and it is also a diagonal matrix. If $n = 1$ then it means V has only one vector say v_1, in the basis and the matrix of f will just be a scalar $f(v_1, v_1) \neq 0$ which is a 1×1 matrix and it is trivially a diagonal matrix. Let U be a subspace of V generated by v_1. We can assume that $f(v_1, v_1) = 0$ because if $f(v_1, v_1) = 0$ then it will follow that $f = 0$ on U and it is the same as the first case.

Let $W = \{v \varepsilon V | f(v_1, v) = 0\}$ W is a subspace of V. We wish to show that $V = U \oplus W$ and to show this, we show

$$U \cap W = \{o\}$$

and

$$V = U + W.$$

We first show $U \cap W = \{o\}$ Suppose $u \varepsilon U \cap W$, then $u \varepsilon U$ and $u \varepsilon W$. Since v_1 genenerates U then $u = cv_1$ for $c \varepsilon K$.

Also since $u \varepsilon U$ and $u \varepsilon W$ then $f(u, u) = 0$ hence

$$0 = f(u_1 u) = f(cv_1, cv_1) = c^2 f(v_1, v_1))$$

but $f(v_1, v_1) \neq 0$ therefore $c^2 = 0$ which implies $c = 0$. This shows that $u = cv_1 = 0.v_1 = 0$ hence $U \cap W = \{o\}$.

Next, we show $V = U + V$. Let $v \varepsilon V$ and set $w = v - \frac{f(v_1, v)}{f(v_1, v_1)} v_1$

Notice that

$$\frac{f(v_1, v)}{f(v_1, v_1)} v_1 \varepsilon U.$$

We only need to show that the element w, as defined above is an element of W. i.e. we need to show that $f(v_1, w) = 0$.

$$f(v_1, w) = f\left(v_1, v - \frac{f(v_1, v)}{f(v_1, v_1)} v_1\right)$$

$$= f(v_1, v) - \frac{f(v_1, v)}{f(v_1, v_1)} (f(v_1, v_1)) = f(v_1, v) - f(v_1, v) = 0$$

hence $f(v_1, w) = 0$ and so $w \varepsilon W$. This shows that

$$v = \left(\frac{f(v_1, v)}{f(v_1, v_1)}\right) v_1 + w$$

hence $V = U + W$ and since $U \cap W = \{o\}$ it then follows that

$$V = U \oplus W.$$

The dimension of $W = n - 1$. By the induction principle, we then assume that the theorem is true for f restricted to a space of dimension less than n, namely it is true for f restricted to W. It implies that there exists a basis $\{u_2, u_3, \ldots, v_n\}$ of W such that the matrix representing f on W is a diagonal matrix i.e. $f(v_i, v_j) = 0$ when $i = j$.

By definition of W, $f(v_1, v_j) = 0$ for $j = 2, \ldots, n$ hence the basis $\{v_1, v_2, \ldots, v_n\}$ of V is such that $f(v_i, v_j) = 0$ for $i = j$. This shows that the matrix $A = (a_{ij}) = (f(v_i, v_j))$ which represents f in the basis $\{v_1, v_2, \ldots, v_n\}$ is a diagonal matrix.

Given a symmetric matrix A, by the preceding theorem, this matrix is congruent to a diagonal matrix. We will introduce a method, using elementary row and column operations to get the diagonal matrix which is congruent to matrix A. Let $A = (a_{ij})$ $i, j = 1, 2, \ldots, n$. There are three cases that can happen to the entries of this matrix.

First case

If $a_{11} \neq 0$ then we apply the row operation

$$R_i \to -a_{i1}R_1 + a_{11}R_i, i = 2, 3, \ldots, n.$$

We then apply the corresponding column operations

$$c_i \to a_{i1}c_1 + a_{11}c_i$$

to reduce A to the form

$$\begin{pmatrix} a_{11} & 0 \\ 0 & B \end{pmatrix}$$

Second case

If $a_{11} = 0$ but $a_{ii} \neq 0$ for some $i > 1$, we apply interchange of rows $R_1 \leftrightarrow R_i$ and then we interchange the corresponding column entries $c_1 \leftrightarrow c_i$ so as to bring the a_{ii} to the first diagonal position. This then reduces the matrix to the first case.

Third case:
If the diagonal entries of A are all zeros i.e. $a_{ii} = 0$ for $i = 1, 2, \ldots, n$ then choose $a_{ij} \neq 0$ for some i,j and apply row operation $R_i \to R_j + R_i$ and then corresponding column operation $c_i \to c_j + c_i$ so as to bring $2a_{ij} \neq 0$ into the ith diagonal. This reduces the matrix to the cases above, we finally reduce A to the form, $\begin{pmatrix} a_{11} & 0 \\ 0 & B \end{pmatrix}$ where B is a symmetric matrix of dimension less than that of A. By induction hypothesis B is brought to diagonal form by elementary row and column operations hence A is finally brought into diagonal form.

Example 8.2.1

Let $A = \begin{pmatrix} 1 & -2 & 3 \\ -2 & 6 & -9 \\ 3 & -9 & 4 \end{pmatrix}$ Find a non-singular matrix P such that $P^t A P$ is diagonal. Find the diagonal form.

Solution

Since A is a symmetric matrix, then by the preceding theorem, it is diagonalizable. We then use sequence of elementary row and column operations to obtain the diagonal form. These same row operations are used on the identity matrix I to give P^t.

$$(A, I) = \begin{pmatrix} 1 & -2 & 3 & : & 1 & 0 & 0 \\ -2 & 6 & -9 & : & 0 & 1 & 0 \\ 3 & -9 & 4 & : & 0 & 0 & 1 \end{pmatrix}$$

applying

$$\begin{pmatrix} R_2 \to 2R_1 + R_2 \\ R_2 \to -3R_1 + R_3 \end{pmatrix} \text{we have} \begin{pmatrix} 1 & -2 & 3 & : & 1 & 0 & 0 \\ 0 & 2 & -3 & : & 2 & 1 & 0 \\ 0 & -3 & -5 & : & -3 & 0 & 1 \end{pmatrix}$$

corresponding column operations give

$$C_2 \to 2C_1 + C_2$$
$$C_3 \to 3C_1 + C_3$$

we then have

$$\begin{pmatrix} 1 & 0 & 0 & : & 1 & 0 & 0 \\ 0 & 2 & -3 & : & 2 & 1 & 0 \\ 0 & -3 & -5 & : & -3 & 0 & 1 \end{pmatrix}$$

$R_3 \to 3R_2 + 2R_3$ we then have

$$\begin{pmatrix} 1 & 0 & 0 & : & 1 & 0 & 0 \\ 0 & 2 & -3 & : & 2 & 1 & 0 \\ 0 & 0 & -19 & : & 0 & 3 & 2 \end{pmatrix}$$

the corresponding column operation gives $c_3 \to -3c_2 + 2c_3$ we then have

$$\begin{pmatrix} 1 & 0 & 0 & : & 1 & 0 & 0 \\ 0 & 2 & 0 & : & 2 & 1 & 0 \\ 0 & 0 & -38 & : & 0 & 3 & 2 \end{pmatrix}$$

$$P = \begin{pmatrix} 120 \\ 013 \\ 002 \end{pmatrix}, P^t = \begin{pmatrix} 100 \\ 210 \\ 032 \end{pmatrix}$$

and

$$P^t AP = \begin{pmatrix} 100 \\ 210 \\ 032 \end{pmatrix} \begin{pmatrix} 1 & -2 & 3 \\ -2 & 6 & -9 \\ 3 & -9 & 4 \end{pmatrix} \begin{pmatrix} 100 \\ 210 \\ 032 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -38 \end{pmatrix}$$

8.3 Quadratic Forms

Definition 8.3.1

Let V be a finite dimensional vector space over a field K and $f: V \times V \to K$ a symmetric bilinear form , the quadratic form associated with f is a function $q: V \to K$ defined by $q(u) = f(u, u)$.

Given a quadratic form q we can recover the symmetric form f associated with q by the identity

$$f(u, v) = \frac{1}{4}q(u+v) - \frac{1}{4}q(u-v)$$

This can be seen as follows

$$\frac{1}{4}q(u+V) - \frac{1}{4}q(u-v) = \frac{1}{4}f(u+v, u+v) - \frac{1}{4}f(u-v, u-v)$$

$$= \alpha\frac{1}{4}\{f(u, +v) + (v, u+v)\} - \frac{1}{4}\{f(u, u-v) - f(v, u-v\}$$

$$= \frac{1}{4}\{f(u,u) + f(u,v) + f(v,u) + f(v,v)\} - \frac{1}{4}\{f(u,u) - f(u,v) - f(v,u) + f(v,v)\}$$

$$= \frac{1}{4}f(u,u) + \frac{1}{4}f(u,v) + \frac{1}{4}f(v,u) + \frac{1}{4}f(v,v)$$

$$- \frac{1}{4}f(u,u) + \frac{1}{4}f(u,v) + \frac{1}{4}f(v,u) - \frac{1}{4}f(v,v), \text{ since } f(u,v) = f(v,u)$$

$$= \frac{1}{2}f(u,v) + \frac{1}{2}f(u,v) = f(u,v)$$

hence

$$f(u, v) = \frac{1}{4}q(u+v) - \frac{1}{4}q(u-v)$$

Example 8.3.1

Consider the bilinear form $f: R^n \times R^n \to R$ defined by

$$f(u, v) = \; = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$$

where $u = u_1, u_2, \ldots, u_n$, $v = v_1, v_2, \ldots, v_n$. Then

$$q(u) = f(u, u) = u_1 u_1 + u_2 \cdot u_2 + \ldots + u_n \cdot u_n = u_1^2 + u_2^2 + \ldots + u_n^2$$

hence $q(u) = \|u\|^2$. Suppose a symmetric bilinear form f is represented by a symmetric matrix $A = (a_{ij})$ then $f : R^n \times R^n \to R$ is defined by $f(X, Y) = X^t A Y$ where $X = (x_1, x_2, \ldots, x_n)$, $Y = (y_1, y_2, \ldots, y_n)$ then the associated quadratic form q is given by

$$q(x) = f(x, x) = X^t A X$$

$$= (x_1, x_2, \ldots, x_n) \begin{pmatrix} a_{11} \, a_{12} \cdots a_{1n} \\ a_{21} \, a_{22} \cdots a_{2n} \\ \vdots \\ a_{n1} \, a_{n2} \cdots a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \sum_{i,j=i}^{n} a_{ij} x_i x_j = a_{11} x_1^2 + a_{22} x_2^2 + \ldots + a_{nn} x_n^2 + 2 \sum_{i \neq j}^{n} a_{ij} x_i x_j$$

Observe that if A is a diagonal matrix then $a_{ij} = 0$ for $i \neq j$ then the last term in the above expression for q(x) is zero, hence

$$q(x) = a_{11} x_1^2 + a_{22} x_2^2 + \ldots + a_{nn} x_n^2$$

Notice also that the symmetric matrix $A = (a_{ij})$ which represents $q(x) = q(x_1, x_2, \ldots, x_n)$ has its diagonal entry a_{ii} equal to the coefficient of x_i^2 and has the entries a_{ij} and a_{ji} each equal to half of the coefficient of $x_i x_j$.

Example 8.3.2

If q is the quadratic form associated with a symmetric bilinear form f, verify that $f(u, v) = \frac{1}{2} (q(u + v) - q(u) - q(v))$.

Solution

$$\frac{1}{2}\{q(u + v) - q(u) - q(v)\} = \frac{1}{2}\{f(u + v, u + v) - f(u, u) - f(v, v)\}$$

$$= \frac{1}{2}\{f(u, u) + f(u, v) + f(v, u) + f(v, v) - f(u, u) - f(v, v)\}$$

since f is symmetric then f(u, v) = f(v, u) hence the above equals $\frac{1}{2}\{2 f(u, v)\} = f(u, v)$

Example 8.3.3

Find the symmetric matrix which belongs to the following quadratic polynomial, $q(x_1, x_2, x_3) = 2x_1^2 - 8x_1 x_2 + x_2^2 - 16x_1 x_3 + 14x_2 x_3 + 5x_3^2$

Solution

Let $A=(a_{ij})\ i,j=1,2,3$ be the required matrix. The diagonal entries a_{ii} of A, equals the coefficient of x_i^2 hence $a_{11}=2$, $a_{22}=1$ and $a_{33}=5$. The entries a_{ij} and a_{ji} each equals half of the coefficient of x_ix_j. Since the coefficient of x_1x_2 is -8, then it follows that $a_{12}=-4$ and $a_{21}=-4$.

Coefficient of x_1x_3 is -16 hence $a_{13}=-8$ and $a_{31}=-8$

Coefficient of x_2x_3 is 14 hence $a_{23}=7$ and $a_{32}=7$

It then follows that

$$A=\begin{pmatrix} 2 & -4 & -8 \\ -4 & 1 & 7 \\ -8 & 7 & 5 \end{pmatrix}.$$

Example 8. 3.4

Consider the quadratic form $q\colon \mathbb{R}^3\to\mathbb{R}$ given by $q(x_1,x_2,x_3)=x_1^2-x_1x_3+x_2^2$. Compute the associated symmetric bilinear form $f\colon \mathbb{R}^3\times\mathbb{R}^3\to\mathbb{R}$.

Solution

This problem can be solved in two ways. In the first case, one can use the identity

$$f(x,y)=\frac{1}{2}\{q(x+y)-q(x)-q(y)\}.$$

In the second case, we can compute the matrix A belonging to the quadratic polynomial and use this matrix to compute the bilinear function. As in example 8.3.3 the matrix A belonging to the quadratic form is

$$A=\begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & 0 \end{pmatrix}$$

then

$$f(x,y)=(x_1,x_2,x_3)\begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}=\left(x_1-\frac{1}{2}x_3,\ x_2,\ -\frac{1}{2}x_1\right)\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$=x_1y_1-\frac{1}{2}x_3y_1+x_2y_2-\frac{1}{2}x_1y_3$$

hence the bilinear form associated with the quadratic form is

$$f(x,y)=x_1y_1-\frac{1}{2}x_3y_1+x_2y_2-\frac{1}{2}x_1y_3$$

where $x=(x_1,x_2,x_3)$ and $y=(y_1,y_2,y_3)$.

Let f a symmetric bilinear form on a vector space V over the set R of real numbers, by theorem 8.2.5, f can be represented by a diagonal matrix. If the number of positive entries in the diagonal is represented by P and the number of negative entries is represented by N. The difference S = P - N is called the signature of f. For instance in example 3.4, the matrix

$$A = \begin{pmatrix} 1 & -2 & 3 \\ -2 & 6 & -9 \\ 3 & -9 & 4 \end{pmatrix}$$

is a symmetric matrix and the diagonal form is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -35 \end{pmatrix}$$

In this diagonal form, P = 2 and N = 1 hence if the bilinear form represented by A is f, then signature of f = 2 - 1 = 1.

8.4 Hermitian Forms

Definition 8.4.1

Let V be a finite dimensional vector space over the field C of complex numbers. The form $f: V \times V \to C$ is called Hermitian form if

(i) $f(au + bw, v) = af(u, v) + bf(w, v)$.

(ii) $f(u, v) = \overline{f(v,u)}$ where $u, v, w, \varepsilon V$ and $a, b, \varepsilon C$ from (i) and (ii) above, we have

(iii) $f(u, cv + dw) = \overline{f(cv+dw,u)} = c\overline{f(v,u) + df(w,u)}$

$$= \overline{cf(v,u)} + \overline{(df(w,u)} = \bar{c}\,\overline{f(v,u)} + \bar{d}\, f\overline{(w,u)} = \bar{c}\, f(u, v) + \bar{d}\, f(u, w).$$

Definition (i) shows that an Hermitian form is linear in the first coordinate while it is conjugate linear in the second coordinate by (iii). Note that if z is a complex number such that $z = a + ib$ and $\bar{z} = z$ then $a + ib = a - ib$. This implies $2ib = 0$, hence $b = 0$.

This means that z = a. This shows that any complex number z which is equal to its conjugate is actually a real number.

By (ii) of definition 8.4.1, $f(u, u) = \overline{f(u,u)}$, then by the above remark, it follows that f(u, u) is a real number for every $u \varepsilon V$. Suppose V is an n-dimensional vector space over the complex numbers C and f an Hermitian form on V. Let $\{e_1, e_2, \ldots, e_n\}$ be a basis for V. A matrix $B = (b_{ij})$ where $b_i = f(e_j ej)$ is called the matrix representation of f in the basis $\{e_i\}$.

Definition 8.4.2

Let A be an n × n matrix over the field C of complex numbers, if $A = (a_{ij})$ then $\overline{A} = (\overline{a_{ij}})$ which is a matrix obtained from A by taking the complex conjugate of every entry of A. The matrix A is said to be Hermitian if $\overline{A}^t = A$ i.e. if $a_{ij} = \overline{a_{ji}}$ for each entry a_{ij} of A.

If $B = (b_{ij}) = (f(e_i, e_j))$ is the matrix representation of the Hermitian form f, then $b_{ij} = f(e_i, e_j) = \overline{f(e_j, e_i)} = \overline{b_{ji}}$ hence B itself is an Hermitian matrix. In particular, since $b_{ii} = f(e_i, e_i) = \overline{fe_i, e_i)} = \overline{b_{ii}}$, then it follows that the diagonal entries b_{ii} of B are real numbers. We now state the complex analog of theorem 8.2.5

Theorem 8.4.1
Let f be an Hermitian form on a vector space V over the field C of complex numbers. There exists a basis $\{e_1, e_2, .., e_n\}$ of V in which f is represented by a diagonal matrix. The proof is similar to the proof of theorem 8.4.1

Example 8.4.1

$$\text{Let } H = \begin{pmatrix} 1 & i & 2+i \\ -i & 2 & 1-i \\ 2-i & 1+i & 2 \end{pmatrix}$$

Since $\overline{H'} = H$ then H is an Hermitian matrix. Find a non-singular matrix P such that $P'H\overline{P}$ is diagonal.

Solution

For an hermitian matrix if we have the following elementary row operations

(1) $R_i \leftrightarrow R_j$

(2) $R_i \rightarrow aR_i \quad a \neq 0$

(3) $R_i \rightarrow aR_j + R_i$

the corresponding column operations are

(1) $c_i \leftrightarrow c_j$

(2) $c_i \rightarrow \bar{a}c_i, \quad \bar{a} \neq 0$

(3) $c_i \rightarrow \bar{a}c_j + c_i$

We first start with matrix

$$(H, I) = \begin{pmatrix} 1 & i & 2+i & : & 1 & 0 & 0 \\ -i & 2 & 1-i & : & 0 & 1 & 0 \\ 2-i & i+i & 2 & : & 0 & 0 & 1 \end{pmatrix}$$

We apply the following elementary row operations

$R_2 \rightarrow iR_1 + R_2$

$R_3 \rightarrow (-2+i)R_1 + R_3$

to obtain

$$\begin{pmatrix} 1 & i & 2+i & : & 1 & 0 & 0 \\ 0 & 1 & i & : & i & 1 & 0 \\ 0 & -i & -3 & : & -2+i & 0 & 1 \end{pmatrix}$$

we then apply the corresponding column operations

$$c_2 \to -ic_1 + c_2$$

$$c_3 \to (-2-i)c_1 + c_3$$

to obtain

$$\begin{pmatrix} 1 & 0 & 0 & : & 1 & 0 & 0 \\ 0 & 1 & i & : & i & 1 & 0 \\ 0 & -i & -3 & : & -2+i & 0 & 1 \end{pmatrix}$$

we then apply the elementary row operation $R_3 \to iR_2 + R_3$ to obtain

$$\begin{pmatrix} 1 & 0 & 0 & : & 1 & 0 & 0 \\ 0 & 1 & i & : & i & 1 & 0 \\ 0 & 0 & -4 & : & -3+i & i & 1 \end{pmatrix}$$

we then apply the corresponding column operation $c_3 \to -ic_2 + c_3$ to obtain

$$\begin{pmatrix} 1 & 0 & 0 & : & 1 & 0 & 0 \\ 0 & 1 & 0 & : & i & 1 & 0 \\ 0 & 0 & -4 & : & -3+i & i & 1 \end{pmatrix}$$

hence

$$P = \begin{pmatrix} 1 & i & -3+i \\ 0 & 1 & i \\ 0 & 0 & 1 \end{pmatrix}$$

and the diagonal form

$$P'H\bar{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -4 \end{pmatrix}$$

signature of $H = 2 - 1 = 1$.

8.5 Exercises

1. Let $f: R^2 \times R^2 \to R$ be a function defined by $f(u, v) = u_1v_1 + u_2v_2$ where $u = (u_1, u_2)$ and $v = (v_1, v_2)$
   (a) show that f is bilinear
   (b) Find the matrix A of f in the basis $\alpha = \{\alpha_1 = (1,2), \alpha_2 = (3,4)\}$
   (c) Find the matrix B of f in the basis $\beta = \{\beta_1 = (1,-1), \beta_2 = (1,1)\}$
   (d) Find the transition matrix P from $\{\alpha_i\}$ to $\{\beta_i\}$ and verify that $B = P'AP$.

2. Let V be vector space of all 2×3 matrices over R and let f be the bilinear form on V defined by

$$f(A, B) = \text{trace}\ (A' M B) \text{ where } M = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

Find the matrix of f in the ordered basis $\{E^{11}, E^{12}, E^{13}, E^{21}, E^{22}, E^{23}\}$ where E is the matrix whose only non-zero entry is 1 in the row i and column j.

3. (a) The following expressions define quadratic forms q on R^3. Find the symmetric bilinear form corresponding to each q.

   (i) $q(x_1, x_2, x_3) = x_1x_2 + x_2^2 + 4x_1x_3 + x_3^2$

   (ii) $q(x_1, x_2, x_3) = 3x_1^2 + 4x_1x_2 - x_2^2 + 8x_1x_3 - 6x_2x_3 + x_3^2$

   (b) Find the matrix, in the standard basis and the rank of the symmetric bilinear forms in (a).

4. Let V be a vector space over a field K. Show that the set of all skew-symmetric bilinear forms on V is a subspace of the space $L(V, V, K)$ of all bilinear forms on V.

5. Let A and B be $n \times n$ matrix over the field of complex numbers. B is said to be Hermitian congruent to A if there exists a non-singular matrix P such that $B = \overline{p}^{t} AP$. Show that Hermitian congruence is an equivalence relation.

6. In each of the matrices below, find the non-singular matrix P such that $P^t A P$ is diagonal.

   (i) $A = \begin{pmatrix} 4 & -3 \\ -3 & -7 \end{pmatrix}$ (ii) $A = \begin{pmatrix} 3 & 2 & 4 \\ 2 & -1 & -3 \\ 4 & -3 & 1 \end{pmatrix}$

7. Let $H = \begin{pmatrix} 2 & 2+i & i \\ 2-i & 3 & 3-i \\ -i & 3+i & 1 \end{pmatrix}$. Find the non-singular matrix P such that $P^t H \overline{P}$ is diagonal. Show the diagonal matrix.

8. If V is a finite dimensional vector space over a field F and $\varphi: V \to F$, $\gamma: V \to F$ are linear functionals on V. Show that $f: V \times V \to F$ define by $f(u, v) = \varphi(u)\, \gamma(v)$ is a bilinear form.

9. Let $f: R^2 \times R^2 \to R$ be defined by $f(u,v) = 3u_1 v_2 - 2u_1 v_2 + 4u_2 v_1 - u_2 v_2$ where $u = (u_1, u_2)$ and $v = (v_1, v_2)$

   (i) Find the matrix A representing f in the basis $e = \{e_1 = (1,1), e_2 = (1,2)\}$

   (ii) Find the matrix B representing f in the basis $e' = \{e_{1'} = (1, -1), e_{2'} = (3, 1)$.

   (iii) Find the transition matrix P from e to e'.

10. Let A and B be $n \times n$ matrices over the set C of complex numbers. We say that B Hermitian congruent to A if there exists a non-singular matrix P such that $B = \overline{P}t\, AP$. Show that Hermitian congruency is an equivalent relation.

11. Let $A = \begin{pmatrix} 1 & 2+3i \\ 2-3i & -1 \end{pmatrix}$ then A is hermitian.

Find a non-singular matrix P such that $P^t A \overline{P}$ is diagonal.

12. Let f be a bilinear form on V. Its is a subset of V, we define

$$S^T = \{v \varepsilon V \mid f(v, u) = 0 \text{ for all } u \varepsilon S\}.$$

(i) show that S^T is a subspace of V.

(ii) If $S_1 \subset S_2$, show that $S_2^T \subset S_1^T$.

CHAPTER NINE

INNER PRODUCT SPACES

In this section, we shall consider vector spaces over a field K where K is the set of real numbers or the set of complex numbers. The objective is to study vector spaces in which it makes sense to speak of the "length" of a vector and "angle" between two vectors. The study will be done by studying a certain scalar-valued function on the Cartesian product of the vector space V which is similar to the Hermitian form which we have just discussed.

9.1 DEFINITION AND EXAMPLES

Definition 9.1.1

Let V be a vector space over a field K where K is the set R of real numbers or the set C of complex numbers. An inner product on V is a function from V × V to K which assigns to each pair of vectors u, v ε V a scalar <u, v> in K such that for all u, v, w ε V and a, b ε K

(i) $<au + bw, v> = a<u, v> + b<w, v>$

(ii) $<u, v> = \overline{< v, u >}$

(iii) $<u, u> \geq 0$ and $<u, u> = 0$ if and only if $u = 0$.

As in the Hermitian case conditions (i) and (ii) implies that

$$<u, aw + bv> = \overline{< aw + bv, u >} = \overline{a < w, u > + b < v, u >}$$
$$= \bar{a}\,\overline{< w, u >} + \bar{b}\,\overline{< v, u >} = \bar{a} < u, w > + \bar{b} < u, v >$$

It can therefore be summarised that a function f: V × V → C is an inner product if f is linear in the first coordinate, conjugate linear in the second coordinate and $f(u, u) > 0$ for $u \underline{=} 0$.

Example 9.1.1

Let C be set of complex numbers, we define a map $<,>: c^n \times c^n \rightarrow C$ by

$< u, v > = u_1\overline{v_1} + \overline{u_2}\,\overline{v_2} + \dots + \overline{u_n}\,\overline{v_n}$ where $u = u_1, u_2, \dots, u_n$, $v = v_1, v_2, \dots, v_n$ Show that $<,>$ is an inner product.

Solution

$$<au + bw, v> = (au_1 + bw_1)\overline{v_1} + (au_2 + bw_2)\overline{v_2} + \dots + (au_n + bw_n)\overline{v_n}$$
$$= au_1\overline{v_1} + bw_1\overline{v_1} + au_2\overline{v_2} + bw_2\overline{v_2} + \dots + au_n\overline{v_n} + bw_n\overline{v_n}$$
$$= a(u_1\overline{v_1} + u_2\overline{v_2} + \dots + u_n\overline{v_n}) + b(w_1\overline{v_1} + w_2\overline{v_2} + \dots + w_n\overline{v_n})$$
$$= a<u, v> + b<w, v> = af(u, v) + bf(w, v).$$

$$\overline{< v, u >} = \overline{v_1\overline{u_1} + v_2\overline{u_2}u_2 + \dots + v_n\overline{u_n}}$$

$$= \overline{v_1}u_1 + \overline{v_2}u_2 + \ldots + \overline{v_n}u_n = u_1\overline{v_1} + u_2\overline{v_2} + \ldots + u_n\overline{v_n} = <u, v>$$

$$<u,u> = u_1\overline{u_1} + u_2\overline{u_2} + \ldots + u_n\overline{u_n} = |u_1|^2 + |u_2|^2 + \ldots + |u_n|^2 \geq 0.$$

Hence this function is an inner product.

Remark (1)

Observe that in Definition 9.1.1 (ii) $<u, u> = \overline{<u, u>}$, hence $<u, u>$ is real.

We denote $\| u \| = \sqrt{<u,u>}$, the non-negative real number which is called the norm or length of u.

Remark (2)

If $<u, u> = 1 = \| u \|$ then u is called a unit vector. Hence a vector whose length equals 1 is called a unit vector. However, any non-zero vector can be reduced to a unit vector. If a non-zero vector is reduced to a unit vector, it is said to be normalized. If $u \varepsilon V$ and $u = \underline{0}$ then

$$w = \frac{u}{\| u \|}$$

is a unit vector because $\| w \| = \left\| \frac{u}{\| u \|} \right\| = \frac{\| u \|}{\| u \|} = 1.$

Definition 9.1.2

A vector space over a field K, where K is set R of real numbers or set C of complex numbers, which has a specified inner product is called an inner product space. A finite-dimensional real inner product space is often called Euclidean space and a complex inner product space is often called a unitary space.

Theorem 9.1.1 (Cauch-Schwartz)

Let V be an inner product space, for any vectors $u, v \varepsilon V$. $|<u, v>| \leq \| u \| \| v \|$

Proof:

If $v = 0$ then the theorem is trivially true because both sides of the above inequality equals zero. We can therefore assume that $v \neq 0$. Let $w = u - \frac{<u, v>}{\| v \|^2} v$ since $\| w \|^2 \geq 0$ then

$$0 \leq \| w \|^2 = \left\langle u - \frac{\langle u, v\rangle}{\|v\|^2} v, u - \frac{\langle u, v\rangle}{\|v\|^2} v \right\rangle$$

$$= \langle u, u\rangle - \left\langle u, \frac{\langle u, v\rangle}{\|v\|^2} v \right\rangle, - \left\langle \frac{\langle u, v\rangle}{\|v\|^2} v, u \right\rangle + \left\langle \frac{\langle u, v\rangle}{\|v\|^2} v, \frac{\langle u, v\rangle}{\|v\|^2} v \right\rangle$$

$$= <u, u> - \frac{\overline{<u, v>}}{\| v \|^2} <u, v> - \frac{<u, v>}{\| v \|^2} <v, u> + \frac{<u, v>}{\| v \|^2} \frac{\overline{<u, v>}}{\| v \|^2} <v, v>$$

$$= <u, u> - \frac{\overline{<u, v> <u, v>}}{\| v \|^2} - <u, v> \frac{\overline{<u, v>}}{\| v \|^2} + \frac{<u, v> \overline{<u, v>}}{\| v \|^2} \| v \|^2$$

$$= \| u \|^2 - \frac{\| <u, v> \|^2}{\| v \|^2} - \frac{\| <u, v> \|^2}{\| v \|^2} + \frac{\| <u, v> \|^2}{\| v \|^2}$$

$$= \|u\|^2 - \frac{2\|<u, v>\|^2}{\|v\|^2} + \frac{\|<u, v>\|^2}{\|v\|^2}$$

$$= \|u\|^2 - \frac{\|<u, v>\|^2}{\|v\|^2} = \frac{\|u\|^2\|v\|^2 - \|<u, v>\|^2}{\|v\|^2}$$

hence

$$0 \leq \frac{\|u\|^2\|v\|^2 - |<u, v>|^2}{\|v\|^2}$$

hence

$$\|u\|^2\|v\|^2 - |<u, v>|^2 \geq 0$$

then

$$\|u\|^2\|v\|^2 \geq |<u, v>|^2$$

therefore

$$<u, v> \leq \|u\|\|v\|.$$

Example 9.1.2

Let V be an inner product space. Show that for any vectors u, v ε V, $\|u + v\| \leq \|u\| + \|v\|$

Solution

This is an application of Cauchy-Schwartz inequality.

$$\|u + v\|^2 = <u + v, u + v> = <u, u> + <u, v> + <v, u> + <v, v>$$

$$= <u, u> + <u, v> + \overline{<u, v>} + <v, v>$$

by Cauchy-Schwartz inequality,

$$<u, v> \leq \|u\|\|v\|.$$

Similarly

$$\overline{<u, v>} \leq \overline{\|u\|\|v\|} = \|u\|\|v\|$$

hence

$$<u, u> + <u, v> + \overline{<u, v>} + <v, v> = \|u\|^2 + <u, v> + \overline{<u, v>} + \|v\|^2$$

$$\leq \|u\|^2 + \|u\|\|v\| + \|u\|\|v\| + \|v\|^2$$

$$= \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2$$

We therefore have

$$\|u + v\|^2 \leq (\|u\| + \|v\|)^2$$

hence

$$\|u + v\| \leq \|u\| + \|v\|$$

9.2 Orthogonality and Orthonormal Vectors

Definition 9.2.1

Let V be an inner product space. The vectors u, v ε V are orthogonal if $<u, v> = 0$. In fact if u is orthogonal to v then $<u, v> = 0$. We can also show that u is orthogonal to v implies v is also orthogonal to u because $<v, u> = \overline{<u, v>} = \overline{0} = 0$, hence v is also orthogonal to u.

Since $(0, v) = <0v, v> = 0 <v, v> = 0$ for all vectors $v \varepsilon V$, it then follows that 0 is orthogonal to every vector in V. Actually zero vector is the only vector in V which is orthogonal to every vector in V. If $u \varepsilon V$ such that u is orthogonal to every vector in V, this means that u must be orthogonal to itself hence $<u, u> = 0$ but by definition of inner product, this implies $u = 0$.

Theorem 9.2.1

Let V be an inner product space. If $(v_1, v_2,\dots, v_n)$ are orthogonal non-zero in V, then $(v_1, v_2,\dots, v_n)$ are linearly independent.

Proof:

Suppose $a_1 v_1 + a_2 v_2 \dots + a_n v_n = 0$ we need to show that $a_1 = 0, a_2 = 0, \dots, a_n = 0$

$$0 = <0, v_1> = <a_1 v_1 + a_2 v_2 + \dots + a_n v_n, v_1)$$

$$= a_1 < v_1, v_1 > + a_2 < v_2, v_1 > + \dots + a_n < v_n, v_1 >$$

since $< v_i, v_1 > = 0$ for $i \underline{=} 1$ because the v_is are orthogonal, then the above equation equals

$$a_1 < v_1 v_1 > + a_2.0 + \dots + a_m.0 = a_1 < v_1, v_1 >$$

$a_1 < v_1, v_1 > = 0$ but $< v_1, v_1 > \neq 0$ since v_1 is non-zero vector, hence $a_1 = 0$.

Similarly,

$$0 = <0, v_2> = <a_1 v_1 + a_2 v_2 + \dots + a_n v_n, v_2>$$

$$= a_1 < v_1, v_2 > + a_2 < v_2, v_2 > + \dots + a_n < v_n, v_2 >$$

$$= a_1.0 + a_2 < v_2, v_2 > + \dots + a_n.0 = a_2 < v_2, v_2 >$$

but $< v_2, v_2 > \neq 0$ since $v_2 \neq 0$ then $a_2 = 0$.

Similarly $a_3 = 0, a_4 = 0, \dots, a_n = 0$

This shows that $(v_1, v_2,\dots, v_n)$ are linearly independent.

Definition 9.2.2

Let W be a subspace of an inner product space V. The orthogonal compliment of W, denoted by W- is the set of all vectors in V which are orthogonal to every vector in W.

$$W\text{-} = \{ v \varepsilon V \mid < v, w > = 0 \text{ for all } w \varepsilon W \}.$$

Theorem 9.2.2

Let V be an inner product space, W a subspace of V. W- is a subspace of V.

Proof:

Recall that the zero vector is orthogonal to every vector, in particular $<0, w> = 0$ for all $w \varepsilon W$ hence $0 \varepsilon W$-, therefore $W\text{-} \neq \phi$. If $u, v \varepsilon W$- and $a, b \varepsilon K$ then $<u, w> = 0$ and $<v, w> = 0$ for every $w \varepsilon W$.

$$<au + bv, w> = a<u, w> + b<v, w> = a.0 + b.0 = 0.$$

hence $au + bv \varepsilon W$-. This shows that W- is a subspace of V.

Definition 9.2.3

Let V be an inner product space, vectors $(v_1, v_2, \dots, v_n)$ are said to be orthonormal vectors if they are orthogonal and $\| v_i \| = 1$ for each $i = 1, 2, \dots, n$.

This implies that the vectors $v_1, v_2, \dots, v_n$ are orthonormal if

$$< v_i, v_j > = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

Example 9.2.1

Consider R^n with the standard inner product which is scalar product $< u, v > = u_1v_1 + u_2v_2 + \dots + u_nv_n$.

The standard basis

$e_1 = (1,0,0,\dots,0)$, $e_2 = (0,1,0,\dots,0)$, $e_3 = (0,0,1,0,\dots,0) \dots e_n = (0,0,0,\dots,0,1)$

is an orthonormal basis.

Solution

$$\begin{aligned} < e_1, e_2 > &= < (1,0,0,\dots,0), (0,1,0,\dots,0) > \\ &= 1.0 + 0.1 + 0.0 + \dots + 0.0 = 0 + 0 + \dots + 0 = 0. \end{aligned}$$

In general $< e_i, e_j > = 0$ for $i \neq j$

$$< e_i, e_i > = 1 \quad i = 1, 2, \dots, n.$$

Theorem 9.2.3

Let V be an inner product space and $\{u_1, u_2, \dots, u_r\}$ orthonormal vectors. Let $v \varepsilon V$ be any vector then the vector

$$w = v - < v, u_1 > u_1 - < v, u_2 > u_2 - \dots - < v, u_r > u_r$$

is orthogonal to each of the u_i.

Proof:

To show that w is orthogonal to each u_i, we need to show

$< w, u_i > = 0$.

$< w, u_i > =$

$$\begin{aligned} &= < v, u_1 > - < v, u_1 >< u_1, u_1 > - < v, u_2 >< u_2, u_1 > - \dots - < v, u_r >< u_r, u_1 > \\ &= < v, u_1 > - < v, u_1 > .1 - < v, u_2 > .0 + \dots + - < v, u_r > .0 \\ &= < v, u_1 > - < v, u_1 > = 0 \end{aligned}$$

hence $< w, u_1 > = 0$ and so w is orthogonal to u_1. For $i = 2, \dots, r$, we have

$< w, u_i > =$

$$= < v, u_i > - < v, u_1 >< u_1, u_i > - < v, u_i >< u_i, u_i > - \dots - < v, u_r >< u_r, u_i >$$

$$= \langle v, u_i \rangle - \langle v, u_1 \rangle .0 - \langle v, u_2 \rangle .0 - \ldots - \langle v, u_i \rangle .1 - 0$$

$$= \langle v, u_i \rangle - \langle v, u_i \rangle = 0$$

hence w is orthogonal to u_i for $i = 1, 2, , \ldots, r$.

Let V be a finite-dimensional inner product space, it is possible to get orthonormal vectors in V which form its basis.

Theorem 9.2.4

Let $\{v_1, v_2, \ldots, v_n\}$ be arbitrary basis of a n-dimensional inner product space V then one can construct orthonormal vectors $\{u_1, u_2, \ldots, u_n\}$ which form a basis for V.

Proof:

This construction is known as Gram-schmidt orthogonalization process. We normalise v_1, to have

$$u_1 = \frac{v_1}{\| v_1 \|}$$

Next we set $w_2 = v_2 - \langle v_2, u_1 \rangle u_1$. By theorem 9.2.3 w_2 is orthogonal to u_1. We then normalize w_2 to have

$$u_2 = \frac{w_2}{\| w_2 \|}$$

Since w_2 is orthogonal to u_1 then u_2 is also orthogonal to u_1 and since u_1 and u_2 are unit vectors then it follows that $\{u_1, u_2\}$ are orthonormal vectors.

We now want to construct u_3. Let $w_3 = v_3 - \langle v_3, u_1 \rangle u_1 - \langle v_3, u_2 \rangle u_2$. By theorem 9.2.3, w_3 is orthogonal to u_1 and u_2 and when we normalise w_3, we have

$$u_3 = \frac{w_3}{\| w_3 \|}.$$

It follows that u_3 is a unit vector and is orthogonal to u_1 and u_2 hence $\{u_1, u_2, u_3\}$ are orthonormal vectors.

In general, after obtaining $\{u_1, u_2, \ldots, u_i\}$ which are orthonormal, to construct u_{i+1}, let $w_{i+1} = v_{i+1} - \langle v_{i+1}, u_1 \rangle u_1 - \ldots - \langle v_{i+1}, u_i \rangle u_i$.

By theorem 9.2.3 w_{i+1} is orthogonal to $u_1, u_2, \ldots, u_i$. When we normalise w_{i+1}, we will have

$$u_{i+1} = \frac{w_{i+1}}{\| w_{i+1} \|}.$$

Since w_{i+1} is orthogonal to $u_1, u_2, \ldots, u_i$, then u_{i+1} is unit vector and it is orthogonal to

$(u_1, u_2, \ldots, u_i)$ hence $\{u_1, u_2, \ldots, u_i, u_{i+1}\}$ is an orthonormal set of vectors.

When we continue inductively, we obtain orthonormal set $\{u_1, u_2, \ldots, u_n\}$. By theorem 9.2.2 they are linearly independent hence they form a basis for V.

Example 9.2.2

Consider the basis $v_1 = (3, 0, 4)$, $v_2 = (-1, 0, 7)$ and $v_3 = (2, 9, 11)$ of R^3. Compute orthonormal basis $\{u_1, u_2, u_3\}$ of R^3.

Solution

$$u_1 = \frac{v_1}{\| v_1 |}$$

since $v_1 = (3, 0, 4)$, then

$$\| v_1 \| = \sqrt{3^2 + 0^2 + 4^2} = \sqrt{9 + 16} = \sqrt{25} = 5.$$

Hence

$$u_1 = \frac{3, 0, 4}{5} = \frac{1}{5}(3, 0, 4)$$

$$w_2 = v_2 - < v_2, u_1 > u_1 = (-1,\ 0,\ 7) - <(-1,\ 0,\ 7),\ \frac{1}{5}(3,\ 0,\ 4) > \frac{1}{5}(3,\ 0,\ 4)$$

$$= (-1, 0, 7) - (-1.\frac{3}{5} + 0.0 + 7.\frac{4}{5})\frac{1}{5}(3, 0, 4)$$

$$= (-1, 0, 7) = \frac{25}{5}\frac{1}{5}(3, 0, 4) = (-1,\ 0,\ 7) - (3,\ 0,\ 4) = (-4,\ 0,\ 3)$$

$$u_2 = \frac{w_2}{\| w_2 \|}$$

$$\| w_2 \| = \sqrt{(-4)^2 + 0^2 + 3^2} = \sqrt{25} = 5$$

hence

$$u_2 = \frac{1}{5}(-4, 0, 3).$$

Let

$$w_3 = v_3 - < v_3, u_1 > u_1 - < v_3, u_2 > u_2$$

$$= (2, a, 11) - < 2, 9, 11), \frac{1}{5}(3, 0, 4) > \frac{1}{5}(3, 0, 4) -$$

$$= (2, 9, 11) - 2.\frac{3}{5} + 9.0 + 11.\frac{4}{5})\frac{1}{5}(3, 0, 4) - (2.(\frac{-4}{5} + 9.0 + 11.\frac{3}{5})\frac{1}{5}(-4, 0, 3)$$

$$= (2, 9, 11) - \frac{50}{5}.\frac{1}{5}(3, 0, 4) - \frac{25}{5}.\frac{1}{5}(-4, 0 3)$$

$$= (2, 9, 11) - 2(3, 0, 4) - (-4, 0, 3) = (0, 9, 0)$$

$$u_3 = \frac{w_3}{\| w_3 \|}$$

but

$$\| w_3 \| = \sqrt{0^2+9^2+0^2} = \sqrt{81} = 9$$

hence

$$u_3 = \frac{0,9,0}{9} = (0,1,0).$$

Hence

$$u_1 = \frac{1}{5}(3,0,4)$$

$$u_2 = \frac{1}{5}(-4,0,3)$$

$$u_3 = (0,1,0).$$

9.3 Linear Functionals and Adjoint Linear Operators

Definition 9.3.1

Let V be a vector space over a field K. A function T: V → K which is linear is called a linear functional on V.

Theorem 9.3.1

Let V be a finite dimensional inner product space and let u ε V be a fixed vector in V, we define $f_u : V \to K$ by $f_u(v) = \langle v, u \rangle$ for every v ε V, f_u is a linear functional on V.

Proof:

To show that f_u is a linear functional is to show that f_u is linear. If v, w ε V and a, b ε K then

$$f_u(av+bw) = \langle av+bw, u \rangle = a\langle v, u \rangle + b\langle w, u \rangle = af_u(v) + bf_u(w)$$

hence f_u is linear and so it is a linear functional. When V is a finite-dimensional inner product space, we will now show that any linear functional on V can be defined this way.

Theorem 9.3.2

Let V be a finite dimensional inner product space on a field K where K = set R of real numbers or set C of complex numbers. If T: V → K is a linear functional on V, then there exists uniquely a vector u ε V such that $T(v) = \langle v, u \rangle = f_u(v)$.

Proof:

Since V is finite-dimensional inner product space, let $\{e_1, e_2, \ldots, e_n\}$ be a basis for V. To show that the linear functional T and f_u are equal on the inner product space V, it is enough to show that they are equal on each vector of the basis $\{e_i\}$ Let

$$u = \overline{T(e_1)}e_1 + \overline{T(e_2)}e_2 + \ldots + \overline{T(e_n)}\, e_n$$

$$= \langle e_1, \overline{T(e_1)}e_1 \rangle + \langle e_1, \overline{T(e_2)}e_2 \rangle + \ldots + \langle e_1, \overline{T(e_n)}e_n \rangle$$

$$= T(e_1) < e_1, e_1 > + T(e_2) < e_1, e_2 > + \ldots + T(e_n) < e_1, e_n >$$

$$= T(e_1).1 + T(e_2).0 + \ldots + T(e_n).0.\ T(e_1).$$

Hence

$$T(e_1) = f_u(e_1).$$

Similarly, for $i = 2, \ldots, n$

$f_u(e_i) = < e_i, u > =$

$$= < e_i, \overline{T(e_1)}e_1 > + < e_i, \overline{T(e_2)}e_2 > + .. + < e_i, \overline{T(e_i)}e_i > + .. + < e_i, \overline{T(e_n)}e_n >$$

$$= T(e_1) < e_i, e_1 > + T(e_2) < e_i, e_2 > + \ldots + T(e_i) < e_i, e_i > + \ldots + T(e_n) < e_i, e_n >.$$

$$= T(e_1).0 + T(e_2).0 + \ldots + T(e_i).1 + \ldots + T(e_n).0 = T(e_i)$$

hence

$$f_u(e_i) = T(e_i) \text{ for all } i = 1, 2, \ldots, n$$

This shows that f_u and T are equal on the basis $(e_1, e_2, \ldots, e_n)$ of V hence they are equal on any vector $v \varepsilon V$. This shows that there exists $u \varepsilon V$ such that $T(v) = < v, u >$ for every $v \varepsilon V$. We now show that the vector u is unique. Suppose there exists another vector u' in V such that $T(v) = < v, u' >$, then $T(v) = < v, u' > = < v, u >$, then $< v, u - u' > = 0$ for every $v \varepsilon V$ and in particular it is also true for $v = u - u'$, i.e. $< u - u', u - u' > = 0$. This is true if and only if $u - u' = 0$ and so $u = u'$. Thus u is unique. The preceding theorem is not true if V is not a finite dimensional inner product space.

Theorem 9.3.3

Let T be a linear operator on a finite dimensional inner product space V. There exists a unique linear operator T^* on V such that $< T(u), v > = < u, T^*(v) >$. The unique linear operator T^* for which the above equality holds is called the adjoint operator to T on V. Moreover, if A is the matrix representing T in an orthonormal basis $\{e_i\}$ of V, then the conjugate transpose $\overline{A}^t = A^*$ of A is the matrix representing T^* in the basis $\{e_i\}$.

Proof:

We first define the operator T^* and then show that it is linear. Let $v \varepsilon V$ be an arbitrary but fixed vector in V.

Consider the map $f : V \rightarrow K$ defined by $f(u) = <T(u), v>$, f is a linear functional. This is true because T is a linear operator. For if $u_1, u_2 \varepsilon V$ and $a, b \varepsilon K$ then

$$f(au_1 + bu_2) = < T(au_1 + bu_2), v > = < aT(u_1) + bT(u_2), v >$$

$$= a < T(u_1), v > + b < T(u_2), v > = af(u_1) + bf(u_2)$$

This shows that f is a linear functional and by theorem 9.3.2 there exist a unique vector $v' \varepsilon V$ such that $f(u) = <u, v'>$ i.e.

$$< T(u), v > = < u, v' > \text{ for every } u \, \varepsilon \, V.$$

We define the operator $T^*: V \to V$ by $T^*(v) = v'$. T^* is a unique map because v' is uniquely determined by theorem 9.3.2; T^* is linear because if u, v, w ε V and a, b ε K then

$$< u, T^*(av + bw) > = < T(u), (av + bw) > = < T(u), av > + < T(u), bw >$$

$$= \bar{a} < T(u), v > + \bar{b} < T(u), w > = \bar{a} < u, T^*(v) > + \bar{b} < u, T^*(w) >$$

$$= < u, a T^*(v) > + < u, b T^*(w) > = < u, a T^*(v) + b T^*(w) >$$

This is true for any vector u ε V hence

$$T^*(av + bw) = a T^*(v) + b T^*(w)$$

and so T^* is linear.

If $A = (a_{ij})$ is the matrix representing the linear operator T and $B = (b_{ij})$ is the matrix representing T^*, then

$$a_{ij} = < T(e_j), e_i > \text{ and } b_{ij} = < T^*(e_j), e_i >$$

but

$$b_{ij} = < T^*(e_j), e_i > = \overline{< e_i, T^*(e_j) >} = \bar{a}_{ji}$$

hence $B = \bar{A}^t = A^*$.

The preceding theorem shows that every linear operator T on a finite inner product space V has an adjoint operator T^* and if A is the matrix representing T with respect to a given basis then $\bar{A}^t$ is the matrix representing T^* in the same basis.

Example 9.3.1
Let T be a linear operator on c^4 defined by
T (x, y, z, w) = (x+iz+3iw, 5y-3iz+w, x+(1+i)y+2z+w, 2x+iy-iz+(1+i)w).
Find $T^*(x, y, z, w)$.

Solution

Since c^4 is a 4-dimensional inner product space over C then the preceding theorem applies. The matrix representing T in the usual basis is

$$A = \begin{pmatrix} 1 & 0 & i & 3i \\ 0 & 5 & -3i & 1 \\ 1 & 1+i & 2 & 1 \\ 2 & i & -i & 1+i \end{pmatrix}$$

By theorem 9.3.3, the matrix representing T^* in the usual basis is $\bar{A}^t$ which is

$$\begin{pmatrix} 1 & 0 & 1 & 2 \\ 0 & 5 & 1-i & -i \\ -i & 3i & 2 & i \\ -3i & 1 & 1 & 1-i \end{pmatrix}$$

Hence

$$T^*(x, y, z, w) = (x + z + 2w, 5y + (1 - i)z - iw - ix + 3iy + 2z + iw,$$
$$-3ix + y + z + (1 - i)w)$$

Theorem 9.3.4

If F and T are linear operators on a finite dimensional inner product space V and c ε K then

(i) $(F+T)^* = F^* + T^*$

(ii) $(cT)^* = \bar{c}T^*$

(iii) $(FT)^* = T^*F^*$

(iv) $(T^*)^* = T.$

Proof:

Let u, w ε V then

$$<(F+T)u, w> = <F(u)+T(u), w> = <F(u), w> + <T(u), w>$$
$$= <u, F^*(w)> + <u, T^*(w)> = <u, F^*(w) + T^*(w)>$$
$$= <u, (F^* + T^*)w>$$

then

$$<F+T)(u), w> = <u, (F^* + T^*)(w)>$$

but

$$<(F+T)(u), w> = <u, (F+T)^*(w)>$$

By uniqueness of adjoint operator then $(F+T)^* = F^* + T^*$

(ii) For u, w ε V,

$<(cT)(u), w> = <cT(u), w> = c<T(u), w> = c<u, T^*(w) = <u, \bar{c}T^*(w)$

then

$$<(cT(u), w> = <u, \bar{c}T^*(w)>$$

but by definition of adjoint operator

$$\langle cT(u), w\rangle = \langle u, \bar{c}T*(w)\rangle$$

by uniqueness of adjoint operator

$$\bar{c}T^* = (cT)^*$$

(iii) Also

$$\langle (FT)(u), w\rangle = \langle FT(u), w\rangle = \langle T(u), F*(w)\rangle = \langle u(T*F*)(w)\rangle$$

but by definition of adjoint operator,

$$\langle (FT), w\rangle = \langle u, (FT)^*(w)\rangle$$

hence by uniqueness of adjoint,

(iv) $= \overline{< w, T^*(u) >} = \overline{< T(w), u >} = ..$

hence

$$\langle T^*(u), w\rangle = \langle u, T(w)\rangle$$

but by definition of adjoint

$$\langle T^*(u), w\rangle = \langle u, (T^*)^*(w)\rangle$$

hence

$$(T^*)^* = T.$$

Example 9.3.3

Let T be a linear operator on an inner product space V. Suppose the adjoint operator T^* of T is such that $T^* = -T$, if λ is an eigenvalue of T, Show that λ is purely imaginary complex number.

Proof:

If $\lambda \varepsilon$ K is an eigenvalue of the linear operator T to show that λ is imaginary we need to show that $\lambda = -\bar{\lambda}$.

For λ to be purely imaginary, if $\lambda = a + ib$ then a must be zero. If $\lambda = -\bar{\lambda}$ then

$$a + ib = -(a - ib) = -a + ib$$

hence

$$a + ib + a - ib = 0$$

$$2a = 0$$

which implies $a = 0$ and so $\lambda = ib$ This shows that λ is purely imaginary. To show that λ is purely imaginary, we need to show that $\lambda = -\bar{\lambda}$. If u, ε V where $u \underline{=} 0$, then

$$\lambda < u, u > = < \lambda u, u > = < T(u), u > = < u, T^*(u) > = < u, -T(u) >$$

$$= < u, -\lambda(u) > = \bar{\lambda} < u, u >$$

This shows that $\lambda < u, u > = -\bar{\lambda} < u, u >$,since $u \underline{=} 0$, then $<u, u> \underline{=} 0$ hence $\lambda = -\bar{\lambda}$ and λ is purely imaginary.

9.4 Orthogonal and Unitary Operators

Definition 9.4.1

Let T be a linear operator on a finite dimensional inner product space V over a field K. If the adjoint T^* of T is such that $T^* = T^{-1}$ then we say that T is an orthogonal operator if K = set R of real numbers, and it is called a unitary operator if K = set C of complex numbers. Notice that $T^* = T^{-1}$ implies $T^* T = I$.

Theorem 9.4.1

Let T be an orthogonal linear operator on a finite dimensional inner product space V, then

(i) T preserves inner product i.e. for u, w ε V, $\langle T(u), T(w)\rangle = \langle u, w\rangle$

(ii) T preserves length of vectors i.e. for u, ε V, $\| T(u) \| = \| u \|$

Proof:

Since T is an orthogonal linear operator, then $T^* = T^{-1}$ hence

(i) $\left\langle T(u), T(w)\right\rangle = \left\langle u, T^*T(w)\right\rangle = \left\langle u, T^{-1}T(w)\right\rangle = \left\langle u, I(w)\right\rangle = \left\langle u, w\right\rangle$ $\langle u, I(w)\rangle$

hence T preserves inner products.

(ii) $\| T(u) \|^2 = \left\langle T(u), T(u)\right\rangle = \langle u, T^*T(u)\rangle$

$$= \langle u, T^{-1}T(u)\rangle = \langle u, I(u)\rangle = \langle u, u\rangle = \| u \|^2$$

hence

$$\| T(u) \| = \| u \|$$

and so T preserves length of vectors.

Suppose T is a unitary operator on a finite dimensional inner product space V and let A be a matrix representing T relative to an orthonormal basis then by theorem 9.3.3 $A^* = \overline{A}^t$ is the matrix that represents T^* in the same orthonormal basis. Since T is unitary then $T^* = T^{-1}$ this implies that the matrix representing T^* which is A^* will be A^{-1}. This motivates the following definition.

Definition 9.4.2

A matrix over the set C of complex numbers is called a unitary matrix if $A^* = A^{-1}$. Equivalently A is unitary if $A^*A = AA^* = I$.

Notice that if A is a matrix over the set R of real numbers then $\overline{A} = A$ hence $A^* = \overline{A}^t = A^t$. This fact motivates the next definition.

Definition 9.4.3

Let A be a matrix over the set R of real numbers. A is called orthogonal if $A^t = A^{-1}$ or equivalently if $A^tA = AA^t = I$.

Theorem 9.4.2

Let A be a unitary (orthogonal) matrix then the rows of A form set of orthonormal vectors and the columns of A form set of orthonormal vectors.

Proof:

Since A is a unitary matrix then $AA^* = A^*A = I$.

If $A = (a_{ij})$ then $A^* = (b_{ij}) = (\overline{a_{ji}})$

If $AA^* = (c_{ij})$

where

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = \sum_{k=1}^{n} a_{ik} \overline{a_{jk}} = a_{i1}\overline{a_{j1}} + a_{i2}a_{j2} + \ldots + a_{in}a_{jn}$$

where R_i = ith row and R_j = jth row

It then follows that $A A^* = (c_{ij}) = I$ if and only if

$$c_{ij} = \begin{cases} 1, & \text{for } i = j \\ 0, & \text{for } i \neq j \end{cases}$$

i.e.

$$\langle R_i, R_j \rangle = \begin{cases} 1, & \text{for } i = j \\ 0, & \text{for } i \neq j \end{cases}$$

This is exactly saying that the rows R_i are orthonormal.

Similarly, if $A^* A = (d_{ij})$ where

$$d_{ij} = \sum_{k=1}^{n} b_{ik} a_{kj} = \sum_{k=1}^{n} \overline{a_{ki}} a_{kj} = \Sigma a_{kj} \overline{a_{ki}}$$

$$= a_{1j}\overline{a_{1i}} + a_{2j}\overline{a_{2i}} + \ldots + a_{nj}\overline{a_{ni}} = \langle (a_{1j}, a_{2j}, \ldots, a_{nj}), (a_{1i}, a_{2i}, \ldots, a_{ni}) \rangle = \langle c_j, c_i \rangle$$

where c_i denotes ith column and c_j denotes jth column. It follow that $A^* A = (d_{ij}) = I$ if and only if

$$d_{ij} = \begin{cases} 1, & \text{for } i = j \\ 0, & \text{for } i \neq j \end{cases}$$

i.e.

$$= \begin{cases} 1, & \text{for } i = j \\ 0, & \text{for } i \neq j \end{cases}$$

This simply means that the column vectors c_i are orthonormal.

Example 9.4.1

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a unitary matrix then $A^* = \overline{A^t} = \begin{pmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{pmatrix}$

and $A A^* = I$

$$\therefore A A^* = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{pmatrix} = \begin{pmatrix} a\overline{a} + b\overline{b} & a\overline{c} + b\overline{d} \\ c\overline{a} + d\overline{b} & c\overline{c} + d\overline{d} \end{pmatrix}$$

$$= \begin{pmatrix} |a|^2 + |b|^2 & a\overline{c} + b\overline{d} \\ c\overline{a} + d\overline{b} & |c|^2 + |d|^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Thus

$$|a|^2+|b|^2=1$$
$$a\bar{c}+b\bar{d}=0$$
$$c\bar{a}+d\bar{b}=0$$
$$|c|^2+|d|^2=1.$$

This shows that the rows of A, $R_1=(a,b)$ and $R_2=(c,d)$ form m orthonormal vectors. Similarly. $A^*A=I$ shows that the columns of A, $c_1=(a,c), c_2=(b,d)$ form orthonormal vectors.

Example 9.4.2

Consider the matrix

$$A=\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The rows of A are orthonormal and the columns are also orthonormal hence A is an orthogonal matrix.

Theorem 9.4.3

Let V be a finite dimensional inner product space. The transition matrix between any two orthonormal basis is a unitary matrix. Also if A = a_{ij}) is a unitary matrix and $e=\{e_1,e_2,\dots,e_n\}$ is orthonormal basis for V then Ae also form an orthonormal basis.

Proof:

Let $\{e_1,e_2,\dots,e_n\}$ be an orthonormal basis while $\{f_1,f_2,\dots,f_n\}$ is another orthonormal basis. Suppose B^t is the transition matrix from $\{e_i\}$ to $\{f_i\}$ then $B=(b_{ij})$ is the matrix of coefficients, when each f_i is written as a linear combination of the e_i's. i.e.

$$f_1=b_{11}e_1+b_{12}e_2+\dots+b_{1n}e_n$$
$$f_2=b_{21}e_1+b_{22}e_2+\dots+b_{2n}e_n$$
$$\vdots$$
$$f_i=b_{i1_{e1}}+b_{i2}e_2+\dots+b_{in}e_n$$
$$\vdots$$
$$f_n=b_{n1}e_1+b_{n2}e_2+\dots+b_{nn}e_{en}$$

hence,

$$f_i=b_{i1}e_1+b_{i2}e_2+\dots+b_{in}e_n$$
$$f_j=b_{ji}e_1+b_2e_2+\dots+b_{jn}e_n$$

since the $f_{i,}s$ are orthonormal basis then

$$= \langle f_i, f_j \rangle = \begin{cases} 1, & \text{for } i=j \\ 0, & \text{for } i \neq j \end{cases}$$

But

$$\langle f_i, f_j \rangle = \langle (b_{i1}e_1 + \ldots + b_{in}), (b_{j1}e_1 + \ldots + b_{jn}) \rangle$$

Since $(e_1, e_2, \ldots, e_n)$ are orthonormal basis

$$\langle e_i, e_j \rangle = \begin{cases} 1, & \text{for } i=j \\ 0, & \text{for } i \neq j \end{cases}$$

then we have

$$\langle f_i, f_j \rangle = b_{i1}\overline{b_{j1}} + b_{i2}\overline{b_{j2}} + \ldots + b_{in}\overline{b_{jn}}$$

This means that

$$b_{i1}\overline{b_{j1}} + b_{i2}\overline{b_{i2}} + \ldots + b_{in}\overline{b_{jn}} = \begin{cases} 1, & \text{for } i=j \\ 0, & \text{for } i \neq j \end{cases}$$

but

$$B = \begin{pmatrix} b_{11}\, b_{12} \cdots b_{1n} \\ b_{21}\, b_{22} \cdots b_{2n} \\ \cdot \quad \cdot \\ \cdot \quad \cdot \\ \cdot \quad \cdot \\ b_{n1}\, b_{n2} \cdots b_{nn} \end{pmatrix}$$

$$B^t = \begin{pmatrix} b_{11}\, b_{21} \cdots b_{n1} \\ b_{12}\, b_{22} \cdots b_{n2} \\ \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \\ b_{1n}\, b_{2n}\, b_{nn} \end{pmatrix}$$

$$b_{i1}\overline{b_{ji}} + b_{i2}\overline{b_{j2}} + \ldots + b_{in}\overline{b_{jn}} = \begin{cases} 1, & \text{for } i=j \\ 0, & \text{for } i \neq j \end{cases}$$

implies that $B\overline{B}^t = BB^* = I$ hence B is a unitary matrix. Since A is a unitary matrix then $AA^* = I$, i.e.

$$a_{1i}a_{1j} + a_{2i}\overline{a_{2j}} + \ldots + a_{ni}\overline{a_{nj}} = \begin{cases} 1, & \text{for } i=j \\ 0, & \text{for } i \neq j \end{cases}$$

$$Ae = e'$$

where $e' = a_{1i}e_1 + a_{2i}e_2 + \ldots + a_{ni}e_n$ then

$$\left\langle e'_i, e'_j \right\rangle = \left(a_{1i}e_1 + a_{2i}e_2 + \ldots + a_{ni}e_n\right), \left(a_{1j}e_1 + a_{2j}e_2 + \ldots + a_{nj}e_n\right)$$

since $\{e_1, e_2, \ldots e_n\}$ is an orthonormal basis then

$$\left\langle e'_i, e'_j \right\rangle = \begin{cases} 0, & \text{for } i = j \\ 1, & \text{for } i \neq j. \end{cases}$$

hence

$$\left\langle e'_i, e'_j \right\rangle = a_{ii}\overline{a_{ij}} + a_{2i}\overline{a_{2j}} + \ldots + a_{ni}\overline{a_{nj}} = \begin{cases} 0, & \text{for } i = j \\ 1, & \text{for } i \neq j \end{cases}$$

hence $\{e_{1'}, e_{2'}, \ldots e_{i'}, e_{j'} \ldots e_{n'}\}$ is an orthonormal basis.

Let T be a linear operator on a finite dimensional inner product space V. Representing T by a diagonal matrix depends on the eigenvalues and eigenvectors of T and hence on the roots of its characteristic polynomial.

Theorem 9.4.4
Let T be a symmetric operator on a real finite dimensional inner product space V. There exists orthonormal basis of V consisting of eigenvectors of T.

Proof:
We need to show that T can be represented by a diagonal matrix relative to an orthonormal basis. There is an alternative version of this theorem for matrices. Let A be a real symmetric matrix, there exists an orthogonal matrix P such that $B = P^t AP = P^{-1} A P$ is diagonal. The proof of the theorem is by induction on the dimension of the inner product space V. If dim V = 1 then the theorem is trivially true because the matrix of T is a 1 × 1 matrix which is diagonal.

Since T is symmetric then $T = T^t$ and since V is an inner product space on the set R of real numbers then $T = T^*$. If λ is an eigenvalue of T with an eigenvector $v \varepsilon V$ then

$$\lambda < v, v > = < \lambda v, v > = < T(v), v > = < v, T^*(v) >$$

since $T = T^*$ then the above

$$= <v, T(v)> = < v, \lambda(v) > = \overline{\lambda} < v, v >$$

Hence

$$\lambda < v, v > = \overline{\lambda} < v, v >,$$

since $v \underline{=} 0$ then

$$< v, v > \neq 0$$

This implies $\lambda = \overline{\lambda}$ and it shows that λ is a real number.
This proves that T, a symmetric operator, has real eigenvalues; then it means that there exists a non-zero eigenvector v_1 of T. Let W be the space spanned by v_1 and let u_1 be a

unit vector in W. We can take $u_1 = \dfrac{v_1}{\| v_1 \|}$

The restriction $\hat{T}$ of T to W^- is also symmetric, dim $W^- = n-1$ and $V = W \oplus W^-$.

By induction hypothesis, the theorem is true for W^- hence there exists an orthonormal basis $\{u_2, u_3, \ldots, u_n\}$ of W^- consisting of eigenvectors of $\hat{T}$ and hence of T. However $< u_1, u_i > = 0$ for $i = 2,3,\ldots n$ because $u_i \in W^-$ and $u_1 \in W$. Thus $\{u_1, u_2, \ldots, u_n\}$ is an orthonormal set of eigenvectors of T.

Example 9.4.3

Let $A = \begin{pmatrix} 7 & 3 \\ 3 & -1 \end{pmatrix}$ This is a symmetric matrix over the set R of real numbers. We find an orthogonal matrix P such that $P^t A P$ is diagonal. We first need to find the characteristic polynomial f(x) of A.

$$f(x) = \begin{vmatrix} 7-x & 3 \\ 3 & -1-x \end{vmatrix} = (7-x)(-1-x) - 9 = 0$$

which implies $-7 - 7x + x + x^2 - 9 = 0$

$$x^2 - 6x - 16 = 0$$
$$(x-8)(x+2) = 0$$

hence the eigenvalues are $x = 8$ and -2. We need to find the eigenvectors, when $x = 8$ then

$$(A - 8I) = \begin{pmatrix} -1 & 3 \\ 3 & -9 \end{pmatrix}$$

This matrix has rank 1 and so the nullity is 1 and so the eigenspace of $x = 8$ is generated by one vector.

Let the vector by $v = (x, y)$ then

$$\begin{pmatrix} -1 & 3 \\ 3 & -9 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and so

$$-x + 3y = 0$$
$$3x - 9y = 0$$

hence

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

$$\therefore v = (3,1)$$

Also when $x = -2$ then

$$(A + 2I) = \begin{pmatrix} 9 & 3 \\ 3 & 1 \end{pmatrix}.$$

The matrix also has rank 1 and so the nullity is 1. The eigenspace of $x = -2$ is generated by one vector. Let this vector be $u = (x, y)$ then $\begin{pmatrix} 9 & 3 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ This implies

$$3x + y = 0$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$$

hence

$$u = (-1, 3)$$

Notice the v and u are orthogonal vectors. We need to normalize the two vectors to obtain orthonormal basis

$$\{u_1, u_2\}$$

$$u_1 = \frac{v}{\|v\|} = \left(\frac{3}{\sqrt{10}}, \frac{1}{\sqrt{10}}\right)$$

$$u_2 = \frac{u}{\|u\|} = \left(\frac{-1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\right)$$

P will be the matrix whose columns are u_1 and u_2 respectively.

$$P = \begin{pmatrix} \frac{3}{\sqrt{10}} & \frac{-1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{pmatrix}$$

$$P^t A P = \begin{pmatrix} 8 & 0 \\ 0 & -2 \end{pmatrix}$$

An orthogonal linear operator need not be a symmetric linear operator and so may not be represented by a diagonal matrix with respect to an orthonormal basis. We shall now discuss other types of linear operators which are representable by diagonal matrices.

Definition 9.4.4
Let V be a finite dimensional inner product space over the set C of complex numbers. A linear operator T on V is called normal if T commutes with its adjoint i.e. T is normal if $TT^* = T^*T$.

Theorem 9.4.5
Let T be a normal linear operator on a finite-dimensional inner product space V over the field of complex numbers.

There exists orthonormal basis for V such that T can be represented by a diagonal matrix with respect to this basis.

Proof:

The proof of this theorem is by induction on dimension of V. If dim V = 1 then the matrix representing T is 1×1 matrix which is trivially diagonal. Suppose dim V = n> 1.

Since V is a complex inner product space then its characteristic polynomial has at least one root, which implies that T has at least one eigenvalue and hence at least one non-zero eigenvector u. Let W be the subspace of V spanned by u and let u_1 be a unit vector in W. Since u is an eigenvector of T, then $T(u) \varepsilon W$, hence W is invariant under T i.e. $T(W) \subseteq W$. We now show that u is also an eigenvector of T^*. Since u is an eigenvector of T, then

$$T(u) = \lambda u,$$

since T is normal then $(T - \lambda I)$ is also normal. $T(u) = \lambda u$ implies $(T - \lambda I)(u) = 0$ since $(T - \lambda I)$ is normal then $(T - \lambda I)^*(x) = 0$ hence $(T^* - \bar{\lambda} I)(u) = 0$ and so $T^*(u) = \bar{\lambda} u$. This shows that u is also an eigenvector of T^* and so W is also invariant under T^*. Let $W-$ be the orthogonal complement of W. Let $u \varepsilon W-$, if $w \varepsilon W$ then $T(w) \varepsilon W$ and so this means that $T^*(u)$ is orthogonal to every $w \varepsilon W$ hence $T^*(u) \varepsilon W-$. This shows that $W-$ is invariant under T^*. Since W is also invariant under T^*, it follows from the above that $W-$ is invariant under $T^{**} = T$. We therefore see that T restricted to $W-$ is also normal and by induction hypothesis since dim $(W-) = n-1$ then there exists orthonormal basis $\{u_2, u_3, \ldots, u_n\}$ with respect to which T is represented by a diagonal matrix. Since $V = W \oplus W-$ and u_1 is the generator of W then $\{u_1, u_2, \ldots, u_n\}$ is an orthonormal basis and T is represented by a diagonal matrix.

Example 9.4.4

Show that the matrix $A = \begin{pmatrix} 2 & i \\ i & 2 \end{pmatrix}$ is normal. Find the unitary matrix P such that $P^* A P$ is diagonal and write the diagonal form.

Since $A = \begin{pmatrix} 2 & i \\ i & 2 \end{pmatrix}$, then $A^* = \begin{pmatrix} 2 & -i \\ -i & 2 \end{pmatrix}$

$$A A^* = \begin{pmatrix} 2 & i \\ i & 2 \end{pmatrix} \begin{pmatrix} 2 & -i \\ -i & 2 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$$

$$A^* A = \begin{pmatrix} 2 & -i \\ -i & 2 \end{pmatrix} \begin{pmatrix} 2 & i \\ i & 2 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$$

hence $A = \begin{pmatrix} 2 & i \\ i & 2 \end{pmatrix}$ is a normal matrix.

$$|A - xI| = \begin{vmatrix} 2-x & i \\ i & 2-x \end{vmatrix} = (2-x)(2-x) - i^2 = 0$$

$$4 - 4x + x^2 + 1 = 0$$

hence

$$x^2 - 4x + 5 = 0$$

$$x = 4 \pm \frac{\sqrt{16-20}}{2}$$

$$x = \frac{4 \pm 2i}{2} = 2 \pm i$$

$$x = 2 + i \text{ and } 2 - i$$

The eigenvalues of A are 2 + i and 2 - i.

To find the eigenvectors, when x = 2 + i then if the eigenvector is u = (x, y) then

$$(A - (2+i)I)(u) = \begin{pmatrix} -i & i \\ i & -i \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$-ix + iy = 0$$

$$ix - -iy = 0$$

The two equations give

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} i \\ i \end{pmatrix}$$

Similarly when x = 2 - i

$$(A - (2-i)I)(u) = \begin{pmatrix} i & i \\ i & i \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$i + iy = 0$$

$$ix + iy = 0$$

hence $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -i \\ i \end{pmatrix}$ the eigenvectors are (i, i) and (-i, i). We however need to normalize these vectors to have

$$\left(\frac{i}{i\sqrt{2}}, \frac{i}{i\sqrt{2}}\right) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

and

$$\left(\frac{-i}{i\sqrt{2}}, \frac{i}{i\sqrt{2}}\right) = \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

the matrix P which diagonalises A will have its columns to be these orthonormal basis i.e.

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

and

$$P^{*} A P = \begin{pmatrix} 2+i & 0 \\ 0 & 2-i \end{pmatrix}.$$

9.5 EXERCISE

1. Show that the function f: $R^2 \times R^2 \to R$ defined by $f(u, v) = u_1 v_1 - 3u_1 v_2 + 3u_2 v_1 + 4u_2 v_2$ is an inner product where $u = (u_1, u_2)$, $v = (v_1, v_2)$.

2. (a) If V is an inner product space and $u, w \varepsilon V$. Show that if u is orthogonal to w then u is orthogonal to any scalar multiple of w.

   (b) Find a unit vector in R^3 orthogonal to both $v_1 = (1,1,2)$ and $v_2 = (0,1,3)$

3. (a) A subspace W of C^3 is generated by the vectors $u_1 = (0,1+i,2)$ and $u_2 = (i,1,1)$ using Gram-Schmidt process, find the orthonormal basis for W.

   (b) If V is an inner product space over a field K then

   (i) show that $\| cu \| = | c | \| u \|$ where u ε V and c ε K

   (ii) $< u, v > = \frac{1}{4} \| u + v \| - \frac{1}{4} \| u - v \|^2 + \frac{i}{4} \| u + iv \|^2 \frac{i}{4} \| u - iv \|^2$

4. (a) Let T be a linear operator on C^3 defined by

   T (x, y, z) = (4x + (2 - i)y, 4ix - (3 + 2i)y, 3ix + (4 + 5i)y - 3z)

   Find $T^*(x, y, z)$

   (b) A matrix B over the set C of complex numbers is said to be Unitarily equivalent to a matrix A if there exists a unitary matrix P such that $B = P^* A P$ where $P^* = \overline{P}^t$. Show that unitary equivalence of matrices is an equivalence relation.

5. (a) Let G be the set of unitary matrices. If the operation on G is multiplication, show that G is a group.

   (b) If $A = \begin{pmatrix} 211 \\ 121 \\ 112 \end{pmatrix}$, find an orthogonal matrix P for which $P^t A P$ is diagonal and write the diagonal form.

6. (a) Let $A = \begin{pmatrix} 1 & 1 \\ i & 3+2i \end{pmatrix}$

   (i) Show that A is a normal matrix.

   (ii) Find a Unitary matrix P such that $P^* A P$ is diagonal and write the diagonal form.

   (b) Let V be a finite dimensional inner product space, U and W are subspaces of V show that (i) $u^- + W^\perp = (U \cap W)^-$ (ii) $(U + W)^- = U^- \cap W^-$

7. If V is an inner product space and u, v ε V, if u is orthogonal to w then

   (i) show that u is also orthogonal to $< w$ where $\subset$ is a scalar.

   (ii) show that $\| u + v \|^2 = \| u \|^2 + \| v \|^2$ if and only if $<u, v> = 0$.

8. If $H = \begin{pmatrix} 1 & i \\ -i & 2 \end{pmatrix}$ Find the rank and signature of H.

9. If A is an orthogonal matrix and it is invertible, show that A^{-1} is also orthogonal.
10. If A and B are orthogonal show that their product AB is also orthogonal.
11. Find the norm of the following vectors.
    (i) $(2 - 5i, 3 + i, 1 - 2i)$
    (ii) $(1 + i, 0, 2)$.

12. Let $A = \begin{pmatrix} -1 & 3 \\ 3 & 7 \end{pmatrix}$. Find an orthogonal matrix P such that $P^t AP$ is diagonal.

BIBLIOGRAPHY

1. Hoffman K. and Kunze R., *Linear Algebra*
2. Ilori S. And Akinyele O., *Elementary Abstract and Linear Algebra*, Ibadan University Press 1986.
3. Lipschutz S., *Linear Algebra*, Schaum's Outline Series McGraw-Hill Book Company, 1968.
4. Reiner I., *Introduction to Matrix Theory*. Holt, Rinehart and Winston Inc. New York 1971.

INDEX

Adjoint, 74, 77, 78, 82, 101, 142-146, 153
adjoint operator, 143-146
alternating form, 120
annihilator, 111, 113, 114
basis, 19, 22-27, 30, 43-46, 51-55, 57-59, 62, 63, 65, 66, 90-100, 102, 103, 106-114, 116-125, 130-133, 139-144, 147, 149-151, 153-156
bilinear form, 115-121, 124, 127-130, 133
cartesian product, 135
Cauchy-Schwartz, 137
characteristic polynomial, 93, 94, 96, 98, 100-103, 151-153
classical adjoint, 101
column matrix, 32
co-domain, 47
determinant, 67-73, 75, 77, 78, 87, 91
diagonal matrix, 32, 33, 78, 90, 91, 96, 124, 125, 128, 130, 131, 133, 151, 153, 154
dimension, 19, 25, 26, 30, 43-46, 51, 52, 66, 93, 99, 104, 106, 114, 125, 126, 151, 153
direct sum, 28
dot product, 4, 10
dual basis, 106, 107, 111, 113
dual space, 104-106, 108, 110
echelon form, 39-44, 52, 95
echelon matrices, 39
eigenspace, 92-95, 152
eigenvalue, 91-97, 100, 103, 146, 151, 153
eigenvector, 91, 94-96, 103, 151, 153-155
elementary row operation, 39, 40, 132
equality of matrices, 33
equivalence relation, 61, 62, 133, 156
free variables, 89
function, 16, 47, 67, 103, 115, 127, 129, 132, 135, 142, 155
functional, 104-106, 110-114, 142, 143
Gram-Schmidt orthogonalization, 140
hermitian form, 130, 131, 135
identity map, 48
identity matrix, 33, 62, 78, 126
image, 47, 48, 50-53, 66, 109, 122
inner product, 4, 135-140, 142-147, 149, 151, 153, 155, 156
inner product space, 136-140, 142-147, 149, 151, 153, 155, 156
integers, 12
inverse mapping, 49
kernel, 50-53, 66, 92
linear combination, 18-23, 25, 26, 29, 30, 40, 51-53, 57, 60, 62, 67, 68, 75, 95, 149
linear dependence, 41
linear equation, 81, 88

linear functional, 104-106, 110-114, 142, 143
linear independence, 41
linear mapping, 49, 50, 55, 66
linear operator, 54, 55, 57, 59, 60, 62, 65, 66, 90-103, 143, 144, 146, 151, 153, 156
mapping, 47-50, 55, 66, 110, 112
matrices, 30-34, 36, 37, 39, 45, 46, 49, 53, 59-62, 65-67, 70, 78, 83, 89, 98, 99, 104, 119, 120, 123, 133, 151, 153, 156
multiplication of matrices, 34, 36
orthogonal, 4, 7, 10, 89, 137-140, 146, 147, 149, 151-156
orthogonal matrix, 89, 149, 151, 152, 156
orthonormal, 137-141, 143, 147-151, 153-156
quadratic form, 127-129
rank, 43-45, 51-53, 89, 93-95, 120, 121, 123, 133, 152, 156
rational numbers, 12
row, 31, 32, 35, 36, 38-44, 46, 67-73, 75, 78, 125, 126, 131-133
scalar, 1-4, 7, 10, 13, 16, 18, 29-31, 33-35, 38, 39, 49, 50, 67, 71, 72, 78, 87, 88, 91-93, 99, 105, 115, 121, 122, 124, 135, 139, 155, 156
scalar multiplication, 1, 2, 13, 16, 29, 33, 49, 105
signature, 130, 132, 156
similar matrices, 98, 99
subspace, 16-19, 26-30, 43, 45, 46, 50, 51, 53, 91, 92, 111, 113, 114, 122, 124, 133, 134, 138, 153, 156
symmetric bilinear form, 120, 121, 124, 127, 128, 130, 133
symmetric matrix, 45, 103, 120, 121, 125, 126, 128, 130, 151, 152
system of linear equations, 67, 80, 81, 84, 86-90
trace, 64, 65, 104, 105, 133
transition matrix, 57-60, 118, 119, 132, 133, 149
triangular matrix, 90
unit vector, 6, 136, 140, 151, 153, 155
unitary matrix, 147-150, 154, 156
unitary space, 136
usual basis, 26, 30, 52, 94, 96, 97, 119, 144
vector, 1-4, 6, 9-30, 35, 41-46, 49-51, 53, 54, 57-67, 80, 90-93, 95-100, 102-106, 108-116, 118, 120-122, 124, 127, 130, 131, 133, 135-140, 142-144, 151-153, 155
vector space, 4, 12-30, 41, 43-46, 49, 51, 53, 54, 57-59, 61, 62, 65, 66, 90-92, 96-100, 102-106, 108, 110, 111, 113-116, 118, 120, 121, 124, 127, 130, 131, 133, 135, 136, 142
zero matrix, 32, 122, 124